UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE LORENA

HOMAILSON LOPES PASSOS

Planejamento de experimentos no ensino da estatística e probabilidade nas
séries finais do ensino fundamental II

Lorena
2018

HOMAILSON LOPES PASSOS

Planejamento de experimentos no ensino da estatística e probabilidade nas séries finais do ensino fundamental II

Dissertação apresentada à Escola de Engenharia de Lorena da Universidade de São Paulo como requisito para a obtenção do título de Mestre em Ciências do Programa de Mestrado Profissional em Projetos Educacionais de Ciências.

Orientador: Dr. Paulo Atsushi Suzuki

Versão Corrigida

Lorena
2018

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE

Ficha catalográfica elaborada pelo Sistema Automatizado
da Escola de Engenharia de Lorena,
com os dados fornecidos pelo(a) autor(a)

Passos, Homailson Lopes

Planejamento de experimentos no ensino da estatística e probabilidade nas séries finais do ensino fundamental II / Homailson Lopes Passos; orientador Paulo Atsushi Suzuki - Versão Corrigida. Lorena, 2018.
118 p.

Dissertação (Mestrado em Ciências - Programa de Mestrado Profissional em Projetos Educacionais de Ciências) - Escola de Engenharia de Lorena da Universidade de São Paulo. 2018

1. Ensino de estatística e probabilidade. 2. Aprendizagem ativa. 3. Planejamento de experimentos. 4. Teoria clássica dos testes. 5. Teoria de resposta ao item. I. Título. II. Suzuki, Paulo Atsushi, orient.

AGRADECIMENTOS

A Deus por permitir a realização deste sonho.

À minha mãe por sempre incentivar a dedicação aos estudos.

À minha esposa pelo suporte e companheirismo prestados para o término deste trabalho.

Ao meu orientador Dr. Paulo Atsushi Suzuki por toda a paciência, orientação e conselhos despendidos durante o desenvolvimento desta dissertação.

Aos meus alunos pela colaboração no projeto com os aviões de papel.

Aos meus colegas de trabalho e do curso pela ajuda e amizade.

*A matemática é o alfabeto com o qual
Deus escreveu o universo*
Galileu Galilei

RESUMO

PASSOS, H. L. Planejamento de experimentos no ensino da estatística e probabilidade nas séries finais do ensino fundamental II. 2018. 118 p. Dissertação (Mestrado em Ciências) – Escola de Engenharia de Lorena, Universidade de São Paulo, Lorena, 2018.

Este trabalho apresenta uma proposta para o ensino da Estatística e Probabilidade nas séries finais do Ensino Fundamental II. Seu objetivo é mostrar que a metodologia aqui adotada possibilita a aquisição de conceitos estatísticos e probabilísticos, assim como o desenvolvimento de habilidades pessoais e interpessoais. Trata de um projeto fundamentado em Planejamento de Experimentos com respaldo na aprendizagem ativa. Na sequência didática do projeto os alunos realizaram um experimento com aviões de papel no qual tiveram que responder, na prática, a seguinte questão “*Quais alterações podem ser feitas em um modelo de avião de papel para que ele permaneça mais tempo no ar?*”. Para atestar a efetividade da sequência didática, foi construído e validado um Teste de Proficiência em Estatística e Probabilidade (TEPEP) com base nos fundamentos da Psicometria. A análise das características do teste foi feita por meio da Teoria Clássica dos Testes e da Teoria de Resposta ao Item. Foram sujeitos da pesquisa 391 alunos de escolas públicas e particulares da região do Vale do Paraíba, Estado de São Paulo. Desse total, 374 auxiliaram na validação do instrumento e os 17 alunos restantes participaram do projeto. Os resultados deste trabalho mostraram que o uso de Planejamento de Experimentos favoreceu a aprendizagem da Estatística e Probabilidade, desenvolvendo também outras competências. Em relação à validação do TEPEP, concluiu-se que os métodos psicométricos empregados têm grande potencial e devem ser mais explorados. Esta pesquisa apresenta, como produtos finais, a metodologia desenvolvida e o teste de proficiência construído, oferecendo ambos a professores e pesquisadores.

Palavras-chave: Ensino de estatística e probabilidade. Aprendizagem ativa. Planejamento de experimentos. Teoria clássica dos testes. Teoria de resposta ao item.

ABSTRACT

PASSOS, H. L. Designs of experiments in the teaching of statistics and probability for the last years of elementary school. 2018. 118 p. Dissertation (Master of Science) – Escola de Engenharia de Lorena, Universidade de São Paulo, Lorena, 2018.

This work presents a proposal for teaching of Statistics and Probability, for the last years of Elementary School. Its objective is to show that the methodology adopted here allows the acquisition of statistical and probabilistic concepts, as well as the development of personal and interpersonal skills. It's a project with a didactic sequence grounded in Designs of Experiments, supported in active learning. In the didactic sequence of the project the students carried out an experiment with paper airplanes in which they had to answer, in practice, the following question *“What changes can be made to a paper airplane model so that it stays longer in the air?”*. To attest the effectiveness of the didactic sequence, it was developed and validated a Proficiency Test in Statistics and Probability (PTSP), this using Psychometry. The analysis of the characteristics of the test was made through the Classical Test Theory and the Item Response Theory. The research subjects were a total of 391 students from public and private schools in the Vale do Paraíba region, State of São Paulo. Of this total, 17 students participated in the project. The other students (374) assisted in the test validation. The results of this research showed that the use of Design of Experiments favored the learning of Statistic and Probability, also to develop others competences. Regarding the validation of PTSP, it could be concluded that the psychometric methods used have potential and they should be more explored. This research acclaims, as final products, the developed methodology and the Proficiency test validated, both offered to teachers and researchers.

Keywords: Teaching of statistics and probability. Active learning. Designs of experiments. Statistic and probability. Classical test theory. Item response theory.

LISTA DE FIGURAS

Figura 1 – Planos de Experimento para fatores com dois níveis.	26
Figura 2 – Componentes do escore real.	38
Figura 3 – Assimetrias e Simetria.	42
Figura 4 – Fluxograma do processo de validação.	48
Figura 5 – Distribuição das cargas das componentes sobre os itens.	65
Figura 6 – Divisão dos grupos segundo a regra de Kelley, paralelo à distribuição normal padrão.	67
Figura 7 – Cálculo de média aritmética feito pelos alunos.	80
Figura 8 – Rol organizado pelos alunos.	81
Figura 9 – Aplicação da fórmula de Sturges e construção da tabela de frequências pelos alunos.	81
Figura 10 – Cálculo de regra de três simples para porcentagem e ângulo, feito pelos alunos.	82
Figura 11 – Alunos utilizando compasso e transferidor.	83
Figura 12 – Conclusão de um dos grupos do projeto sobre o experimento.	84

LISTA DE GRÁFICOS

Gráfico 1 – Percentual de acertos dos alunos nas questões de Estatística e Probabilidade nos exames do SARESP (anos de 2011 a 2014).....	18
Gráfico 2 – <i>Screeplot</i>	36
Gráfico 3 – Representação da “Regra 27”.	40
Gráfico 4 – Exemplo de Curva Característica do Item (CCI).....	43
Gráfico 5 – Correlação entre o desempenho acadêmico dos alunos e o desempenho no TEPEP.	62
Gráfico 6 – <i>Screeplot</i> em análise paralela.	63
Gráfico 7 – Distribuição normal dos escores dos 374 alunos.	66
Gráfico 8 – Divisão dos grupos inferior e superior.	68
Gráfico 9 – <i>Scatterplot</i> dos dados coletados contra os dados estimados pelo modelo.	70
Gráfico 10 – Curva Característica do TEPEP.	71
Gráfico 11 – Histograma da proficiência em estatística e probabilidade (θ), dos 374 alunos.	71
Gráfico 12 – Resultado do Pré-Teste.	73
Gráfico 13 – Curva Característica do Item 05.	75
Gráfico 14 – Curva Característica do Item 09.	75
Gráfico 15 – Curva Característica do Item 10.	76
Gráfico 16 – Curva Característica do Item 14.	77
Gráfico 17 – Curva Característica do Item 15.	77
Gráfico 18 – Gráfico de colunas construído pelos alunos.	80
Gráfico 19 – Histograma construído pelos alunos.	82
Gráfico 20– Gráfico de setores construído pelos alunos.....	83
Gráfico 21 – Resultados do Pré-teste e Pós-teste.....	84
Gráfico 22 – Escore dos itens no Pré-teste e Pós-teste.	85

LISTA DE TABELAS

Tabela 1 – Guia para o Planejamento de Experimentos.	25
Tabela 2 – Regra para interpretação do coeficiente de correlação.	30
Tabela 3 – Regra de interpretação do coeficiente de correlação.	31
Tabela 4 – Interpretação do índice de discriminação de um item.....	40
Tabela 5 – Resultados de Coeficiente de Validade de Conteúdo (CVC) do julgamento do TEPEP.....	59
Tabela 6 – Desempenho dos alunos.....	61
Tabela 7 – Análise das principais componentes e fatores.	64
Tabela 8 – Variância acumulada segundo a Análise das Componentes Principais (ACP)..	64
Tabela 9 – Cargas das componentes.	65
Tabela 10 – Índices de discriminação dos itens segundo os resultados do Teste T.	69
Tabela 11 – Índices de discriminação dos itens segundo a correlação ponto-bisserial.	69
Tabela 12 – Índices do modelo de dois parâmetros.....	72
Tabela 13 – Escores nos 15 itens dos alunos com desempenho abaixo da média.	74
Tabela 14 – Descritores dos itens.	74

LISTA DE QUADROS

Quadro 1 – Bases metodológicas da sequência didática do Projeto.	51
Quadro 2 – Alterações sugeridas pelos alunos.....	53
Quadro 3 – Desempenho dos grupos superior, médio e inferior, segundo a TCT.....	68

LISTA DE SIGLAS

ACP	Análise das Componentes Principais
AFE	Análise Fatorial Exploratória
CCI	Curva Característica do Item
CVC	Coeficiente de Validade de Conteúdo
SARESP	Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo
TCT	Teoria Clássica dos Testes
TEPEP	Teste de Proficiência em Estatística e Probabilidade
TRI	Teoria de Resposta ao Item

SUMÁRIO

1	INTRODUÇÃO	17
2	OBJETIVOS	20
2.1	GERAL	20
2.2	ESPECÍFICOS	20
3	REVISÃO BIBLIOGRÁFICA	21
3.1	APRENDIZAGEM ATIVA	21
3.2	PLANEJAMENTO DE EXPERIMENTOS	25
3.3	DELINEAMENTO DE PRÉ-TESTE E PÓS-TESTE COM UM GRUPO	27
3.4	VALIDAÇÃO DE INSTRUMENTOS	28
3.5	ANÁLISE FATORIAL EXPLORATÓRIA	33
3.6	TEORIA CLÁSSICA DOS TESTES	38
3.7	TEORIA DE RESPOSTA AO ITEM	43
4	METODOLOGIA	47
4.1	PÚBLICO ALVO	47
4.2	TESTE DE PROFICIÊNCIA EM ESTATÍSTICA E PROBABILIDADE	48
4.3	A ESTATÍSTICA DOS TEMPOS DE VOOS DOS AVIÕES DE PAPEL	51
4.4	FERRAMENTAS TECNOLÓGICAS UTILIZADAS	57
5	RESULTADOS E DISCUSSÃO	59
5.1	VALIDAÇÃO DO TESTE DE PROFICIÊNCIA EM ESTATÍSTICA E PROBABILIDADE	59
5.2	TEORIA CLÁSSICA DOS TESTES	66
5.3	TEORIA DE RESPOSTA AO ITEM	70
5.4	RESULTADOS DO PROJETO	73
6	CONSIDERAÇÕES FINAIS	87
7	PRODUTOS DECORRENTES DESTE TRABALHO	90
8	REFERÊNCIAS	91
	APÊNDICE A – Teste de Proficiência em Estatística e Probabilidade (TEPEP)	99
	APÊNDICE B – Tabela para registro do tempo de voo dos aviões de papel	107
	APÊNDICE C – Roteiro para organização dos grupos durante o experimento	108
	APÊNDICE D – Dimensões e linhas de corte (ARUMAKAN)	109
	APÊNDICE E – Saída do Programa Minitab (Plano do Experimento)	110

APÊNDICE F – Resultados do julgamento dos juízes	111
APÊNDICE G – Curva Característica dos Itens (CCI)	112
APÊNDICE H – Dispositivo lançador de aviões de papel	114
ANEXO A – Modelo de avião Arumakan	115
ANEXO B – Modelo de avião D’Artagnan	116
ANEXO C – Modelo de avião Glider	117
ANEXO D – Modelo de avião Dart	118

1 INTRODUÇÃO

Com o advento da Era da Informação (a partir de 1980), a maior parte das pessoas passou a viver em uma cultura baseada na comunicação digital. Antes disso, as informações eram divulgadas por jornais, revistas e televisão. A difusão da *Internet* e da comunicação móvel, por meio dos *smartphones*, fez com que as informações chegassem muito mais rápido às pessoas. Essas informações, muitas vezes, são apresentadas por intermédio de objetos matemáticos como gráficos e tabelas. Essa nova dinâmica na forma de comunicação entre as pessoas impulsionou ainda mais o uso da Estatística e Probabilidade, tanto para uso comum ou profissional, pois está associado à grande quantidade de informação veiculada pelas mídias (BRASIL, 1998). Desse modo, cada vez mais a população é exposta a dados estatístico-probabilísticos, revelando que a Estatística e a Probabilidade estão intimamente ligadas às relações da sociedade, exigindo, de todos, conhecimento mínimo do tema.

No Brasil, o Ministério da Educação passou a dar maior relevância ao ensino da Estatística e Probabilidade a partir da década de 90. Um dos motivos é porque a disciplina possibilita aos alunos desenvolvimento cognitivo, por meio de modos característicos do pensamento estatístico, observáveis na resolução de situações-problema que demandem coleta, organização, interpretação e apresentação de dados (BRASIL, 1998). Apesar da importância adquirida, os esforços para desenvolver métodos eficientes de avaliação em Estatística e Probabilidade ainda são poucos, comparados ao empenho para situá-la no patamar dos currículos educacionais (GAL; GARFIELD, 1997). Isso desperta o interesse em saber como a disciplina está sendo avaliada e sua atual condição de ensino, especialmente no nível básico, em que seus primeiros conceitos são apresentados.

No Estado de São Paulo, por exemplo, o instrumento utilizado para mensurar o nível de proficiência em Estatística e Probabilidade dos alunos é o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP). Resultados anuais desse indicador trazem pouca informação quanto ao desempenho dos alunos no tema. Buscando uma visão mais ampla sobre o comportamento dos estudantes no assunto, extraiu-se, de cada relatório pedagógico publicado pelo SARESP, nos anos de 2011 a 2014, o percentual de acertos dos alunos do Ensino Fundamental II (EFII) nas questões relacionadas aos conceitos estatísticos, compilados no Gráfico 1 (SÃO PAULO, 2011, 2012, 2013, 2014).

Gráfico 1 – Percentual de acertos dos alunos nas questões de Estatística e Probabilidade nos exames do SARESP (anos de 2011 a 2014).

Fonte: Relatórios pedagógicos do SARESP.

Nota-se que, em geral, comparado à última série do Ensino Fundamental I (5º ano), o rendimento dos alunos das séries seguintes diminui, indicando piora na qualidade do ensino da Estatística e Probabilidade nesse período. Vários podem ser os fatores responsáveis por essa realidade.

Walichinski e Santos Junior (2013) ressaltam que a atenção dada à Estatística e Probabilidade na escola não é tanta quanto na universidade. Garfield e Ahlgren (1988) destacam que os estudantes, em geral, resolvem problemas de Matemática por meio de um “processador de números”, atribuindo valores a uma fórmula computacional ou procedimento sem uma compreensão interna do problema. Para Cobb e Moore (1997) existe uma diferença significativa entre o ensino da Matemática e da Estatística e Probabilidade, o que pode não estar sendo considerado na prática docente. Fatores sociais e emocionais não são descartados (CHACÓN, 2002).

Por conta desses problemas, práticas docentes e pesquisas teóricas sobre o ensino da Estatística e Probabilidade vêm sendo desenvolvidas. Garfield e Ben-zvi (2007), por exemplo, orientam que as aulas foquem no discente, com grupos de três a quatro alunos, para aperfeiçoar a capacidade de resolução de problemas, discutir procedimentos ou analisar dados. Cobb (1992) incentiva a aprendizagem ativa, menos aulas expositivas e mais grupos de resolução de problemas. Cita também, como boa prática, o experimento de George Box (1992), que utilizou um helicóptero de papel para ensinar, de forma ativa, Planejamento de Experimentos a alunos de engenharia. Experimento, esse, que inspirou o

desenvolvimento da metodologia de ensino aqui aplicada, a qual possui Planejamento de Experimentos como parte de sua base.

Pesquisas relacionadas ao uso de Planejamento de Experimentos como metodologia ativa para o ensino da Estatística e Probabilidade estão direcionadas, em geral, para cursos de graduação das engenharias (PENZIAS, 1989; BISGAARD, 1991; LUDLOW; SCHULZ; ERJAVEC, 1995; BOER; CATEN; PAULA, 2014). No Ensino Médio e Fundamental, não foram encontradas publicações sobre o emprego desse tema como estratégia de ensino da disciplina. Esse panorama despertou maior interesse para se investigar a abordagem dessa prática em tais séries escolares.

Portanto, esta pesquisa buscou verificar se o método de ensino aqui empregado teve influência estatística sobre a mudança do nível de proficiência dos alunos em Estatística e Probabilidade. Inclusive, por meio de ferramentas próprias da Psicometria, foi construído e validado um teste para avaliar o grau de conhecimento estatístico e probabilístico dos alunos, antes e depois da aplicação das atividades de ensino.

Em suma, esta dissertação é apresentada na seguinte sequência: Objetivos, Revisão bibliográfica, Metodologia, Resultados e discussão, Considerações finais, Produtos decorrentes deste trabalho e Referências.

2 OBJETIVOS

2.1 GERAL

1) Aprimorar o nível de proficiência em Estatística e Probabilidade dos alunos das séries finais do Ensino Fundamental II (EFII);

2) Validar o Teste de Proficiência em Estatística e Probabilidade (TEPEP), utilizando ferramentas próprias da Psicometria.

2.2 ESPECÍFICOS

Da aplicação da metodologia de ensino desenvolvida:

- Proporcionar aos alunos:
  - a) leitura e interpretação de dados expressos em tabelas e gráficos;
  - b) representação e contagem de casos possíveis em situações combinatórias;
  - c) organização de dados e construção de tabelas e gráficos;
  - d) compreensão dos termos: frequência, frequência relativa, população e amostra;
  - e) realização e distribuição de frequências de uma variável em intervalos de classes;
  - f) obtenção de medidas de tendência central (média, moda e mediana);
  - g) resolução de problemas envolvendo noções de porcentagem;
  - h) desenvolvimento de habilidades pessoais e interpessoais como: trabalho em equipe, liderança, capacidade de argumentação, etc.

Da validação do TEPEP:

- Atingir os critérios de validade necessários para que o teste seja confiável (validade de conteúdo, critério e construto);
- Comprovar a unidimensionalidade do teste;
- Determinar as características do TEPEP, por intermédio da aplicação da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI), realizando um estudo comparativo de seus resultados;
- Utilizar o TEPEP, já validado e caracterizado, na análise dos resultados quantitativos, obtidos a partir aplicação da metodologia de ensino.

3 REVISÃO BIBLIOGRÁFICA

Neste capítulo serão apresentados os referenciais teóricos acerca dos seguintes conceitos: Aprendizagem ativa; Planejamento de Experimentos; Delineamento Pré-Experimental de Pré-Teste e Pós-Teste com um grupo; Psicometria e Validação de Instrumentos.

3.1 APRENDIZAGEM ATIVA

Nas últimas duas décadas, professores e pesquisadores da educação têm demonstrado forte interesse por metodologias ativas de ensino (CROSS, 1987; BONWELL; EISON, 1991; STERN, HUBER, 1997; SILBEMAN, 1998; SANZ, 1999; NIEME, 2002). Essas práticas apresentam-se, geralmente, como alternativas a aulas expositivas, contrapondo a ideia de um aluno passivo no processo de aprendizagem.

Estabelecer um vocabulário absoluto para todos os termos que se referem à aprendizagem ativa é uma tarefa impossível, visto que vários autores trazem diferentes interpretações sobre eles. Há, porém, definições genericamente aceitas (PRINCE, 2004):

- ***Aprendizagem ativa*** é definida como qualquer metodologia de ensino que envolva os estudantes no processo de ensino e aprendizagem, onde estejam realizando atividades que os conduzam a reflexões sobre o que estão fazendo (BONWELL; EISON, 1991). Este conceito é profundamente arraigado à filosofia educacional de John Dewey (1959).
- ***Aprendizagem colaborativa*** tem relação com a metodologia de ensino na qual os estudantes trabalham juntos em pequenos grupos para atingirem uma meta comum, ao invés de aprender solitariamente (GOODSELL et al., 1992).
- ***Aprendizagem cooperativa*** corresponde ao grupo de trabalho estruturado, onde os estudantes buscam cumprir, coletivamente, o mesmo objetivo. Nesse modelo, os alunos trocam informações, compartilham habilidades e avaliam ideias. A principal característica dessa prática é a colaboração dos membros da equipe, contrária a aprendizagem com base na competição (JOHNSON; JOHNSON; SMITH, 1998).

- **Aprendizagem baseada em problemas (ABP)** associa-se a metodologia educacional que introduz problemas relevantes aos alunos no início da sequência didática, provendo motivação e contexto para fluir a aprendizagem. Geralmente trabalha-se com grupos para resolução do problema. É uma metodologia com ênfase na aprendizagem autodirigida (BARROWS; TAMBLYN, 1980).

Nota-se que há uma diversidade de metodologias ativas. Nesta revisão da literatura, são enfatizadas as características da **Aprendizagem colaborativa** (GOODSELL et al., 1992) e da **Aprendizagem ativa** (BONWELL; EISON, 1991), fundamentais para esta pesquisa.

Goodsell et al. (1992), em sua clássica obra “*Collaborative Learning: A Sourcebook for Higher Education*”, apresenta nove artigos discutindo a Aprendizagem colaborativa. No primeiro artigo, “*What is Collaborative Learning?*”, Smith e MacGregor (1992) caracterizam a **Aprendizagem colaborativa** como: situações onde os alunos trabalham em grupos de dois ou mais, mutuamente buscando por compreensão, soluções, significados ou pela criação de um produto. Todos os alunos participam, trabalhando como parceiros ou em pequenos grupos. Problemas, questões ou desafios para criar algo dirigem a atividade. Os autores ressaltam que, embora a Aprendizagem colaborativa esteja um pouco distante das aulas centradas no professor, as aulas colaborativas não pretendem suprimir completamente as aulas expositivas, mas complementá-las. Reforçam que o professor que lida com a Aprendizagem colaborativa deve pensar menos em si como transmissor do conhecimento, passando a se portar como mediador do conhecimento. Eles descrevem cinco premissas para esse tipo de aprendizagem:

- 1) É um processo ativo e construtivo: considera-se o conhecimento prévio dos alunos para se criar um novo saber;
- 2) Depende de um contexto amplo: as atividades devem envolver os alunos e começar com um problema atrativo;
- 3) Os alunos são diversos: aprendem de forma diferente, têm concepções diferentes, possuem experiências diversas, etc.;
- 4) É inerentemente social: os alunos se comunicam, fazem trocas de ideias;

- 5) Tem dimensões afetivas e subjetivas: a aprendizagem cria conexões entre alunos e ideias, entre estudantes e professores. Uns ouvem o ponto de vista dos outros, sendo passíveis de mudança;

Outro ponto importante abordado por Smith e MacGregor (1992), relacionado à aprendizagem colaborativa, é o das três metas para a educação:

- 1) Envolvimento: os alunos são inteiramente ocupados com tarefas extraclases, têm múltiplas responsabilidades em casa, são desconexos entre si ou com a comunidade escolar. A Aprendizagem colaborativa procura criar um envolvimento do aluno em todos os aspectos: social, acadêmico, etc.;
- 2) Cooperação e trabalho em equipe: durante as atividades os alunos acabam observando diferenças de ponto de vista, comportamentos e atitudes. A aprendizagem colaborativa busca desenvolver a capacidade de tolerância, aceitação para todas as vozes do grupo. Cultivar o trabalho em equipe e a habilidade de liderança é meta legítima e valiosa da sala de aula e não extracurricular.
- 3) Responsabilidade cívica: a aprendizagem colaborativa encoraja os estudantes a terem autonomia e a desenvolverem o pensamento crítico e reflexivo, auxiliando na criação do senso democrático.

Bonwell e Eison (1991) dissertam sobre os pontos positivos da *Aprendizagem ativa*, as barreiras para implantá-la, os contrapontos às aulas puramente expositivas e procuram descrever, segundo suas bases literárias, as características gerais desse modelo de aprendizagem. Para os autores, em uma estratégia de Aprendizagem ativa:

- os alunos não são apenas ouvintes, são agentes ativos da aprendizagem;
- dá-se menor ênfase à transmissão de informação e maior atenção ao desenvolvimento das habilidades dos alunos;
- os alunos mentalizam pensamentos de alta ordem (analisar, sintetizar e avaliar);
- os estudantes estão engajados em atividades (ler, escrever, discutir, etc);
- dá-se maior ênfase à exploração das atitudes e valores dos próprios estudantes.

Na descrição, os autores definem a aprendizagem ativa como uma tarefa que “envolva o estudante na atividade que está realizando e pense sobre o que está fazendo” (BONWELL; EISON, 1991, p. 19, tradução nossa). Citam diferentes práticas de ensino caracterizadas como ativas e sugerem que os professores intercalem estratégias alternativas de ensino com o método expositivo, descentralizando o papel de professor. Recomendam que, durante a instrução, os alunos utilizem certo tempo para realizarem atividades onde sejam responsáveis por sua execução, com intervenções mediadas pelo docente.

Buscando tratar sobre estratégias adicionais para promover aprendizagem ativa, os autores descrevem o método de Resolução de Problemas (DEWEY, 1959). É composto por uma sequência de quatro estágios: (1) definir o problema; (2) diagnosticar as possíveis razões do problema; (3) procurar soluções alternativas para o problema; (4) avaliar as alternativas e escolher a solução mais apropriada (BONWELL; EISON, 1991, p. 52).

Outra possibilidade muito interessante, citada ainda no quarto capítulo, é o *Guided Design* (Projeto Dirigido), desenvolvido pela Universidade de West Virginia, em 1970. Muito similar à estratégia descrita no parágrafo anterior, é um processo que formaliza os passos muitas das vezes feitos, inconscientemente, ao tomarmos decisões no dia a dia:

1. Reconhecer um problema e estabelecer uma meta;
2. Reunir informações relevantes;
3. Gerar possíveis soluções para o problema;
4. Listar restrições ao que pode ser feito;
5. Escolher uma solução possível;
6. Analisar a importância dos fatores que precisam ser considerados no desenvolvimento de uma solução detalhada;
7. Criar (sintetizar) uma solução detalhada;
8. Avaliar a solução final;
9. Recomendar um curso de ação (NARDI; WALES, 1985).

3.2 PLANEJAMENTO DE EXPERIMENTOS

Montgomery (2001) define Planejamento de Experimentos como o processo em que se estabelece um plano para um experimento e se coletam dados passíveis de análise estatística, resultando em conclusões válidas e objetivas. O autor resume, conforme a Tabela 1, os passos para a realização do processo.

Tabela 1 – Guia para o Planejamento de Experimentos.

1. Reconhecimento do problema.	}	Planejamento Pré-experimento
2. Escolha dos fatores, níveis e alcance.*		
3. Seleção da variável resposta.*		
4. Escolha do plano do experimento.		
5. Desenvolvimento do experimento.		
6. Análise estatística dos dados.		
7. Conclusões e recomendações.		

* Na prática, os passos 2 e 3 são muitas vezes feitos simultaneamente ou em ordem contrária.

Fonte: Adaptado de Montgomery (2001, p. 14).

Resumo das etapas definidas pelo autor:

- 1. Reconhecimento do problema:** nessa etapa é necessário desenvolver todas as ideias a respeito dos objetivos do experimento, ou seja, a solução do problema. Geralmente, é necessário solicitar sugestões de todos os envolvidos. Por isso, o trabalho em equipe para planejar experimentos é recomendado.
- 2. Escolha dos fatores, níveis e alcance:** devem ser considerados os fatores que mais influenciam no desempenho do experimento, chamados de **fatores potenciais**. Por exemplo, em um experimento onde se deseja descobrir qual o melhor processador a se utilizar em computador para executar determinada tarefa, os fatores potenciais seriam os diferentes processadores utilizados.
- 3. Seleção da variável resposta:** deve ser a variável que realmente provê informação útil sobre o caso estudado. No caso do exemplo citado anteriormente, pode-se dizer que uma das variáveis resposta poderia ser o tempo de conclusão da tarefa.
- 4. Escolha do plano do experimento:** o tipo de plano a ser escolhido depende do tamanho da amostra. Geralmente é realizada através de algum *software* estatístico, onde são fornecidas as informações sobre o experimento e o programa retorna possibilidades de planos para o experimento.

5. **Desenvolvimento do experimento:** ao se realizar o experimento deve-se ter certeza que todos os procedimentos estão sendo realizados de acordo com o planejado. Erros nesse passo serão prejudiciais à validade do experimento.
6. **Análise Estatística dos dados:** métodos estatísticos devem ser utilizados para análise dos dados. *Softwares* estatísticos são empregados para realizar esta tarefa.
7. **Conclusões e recomendações:** após a análise dos dados, são feitas conclusões a respeito dos resultados e recomenda-se um curso de ação. Gráficos facilitam o cumprimento desta etapa.

O plano de experimento a ser seguido (etapa 4) varia conforme o número de fatores e do que se pretende realizar. Para um experimento com uma quantidade grande de fatores, aconselha-se a utilização de planos reduzidos (fracionados). Para experimentos com poucos fatores, pode-se empregar planos completos. Na Figura 1, estão esquematizados os planos de experimento para até 11 fatores e dois níveis. Os níveis compreendem quantidade de possibilidades para cada fator.

Figura 1 – Planos de Experimento para fatores com dois níveis.

Número de Fatores										
Nº de rodadas	2	3	4	5	6	7	8	9	10	11
	2^2	2^{3-1}								
	4	2^3	2^{4-1}	2^{5-2}	2^{6-3}	2^{7-4}				
		8	2^4	2^{5-1}	2^{6-2}	2^{7-3}	2^{8-4}	2^{9-5}	2^{10-6}	2^{11-7}
			16	2^5	2^{6-1}	2^{7-2}	2^{8-3}	2^{9-4}	2^{10-5}	2^{11-6}
				32	2^6	2^{7-1}	2^{8-2}	2^{9-3}	2^{10-4}	2^{11-5}
					64	2^7	2^{8-1}	2^{9-2}	2^{10-3}	2^{11-4}
						128				

	Fatorial completo		Fatorial fracionado de resolução IV
	Fatorial fracionado de resolução V (ou mais)		Fatorial fracionado de resolução III

Fonte: Adaptado de Montgomery (2001, p. 663 – 679).

As informações nas células dizem respeito à forma de planejamento 2^{k-p} . Em que k representa o número de fatores e $(\frac{1}{2})^p = 2^{-p}$ a fração desejada do fatorial total 2^k , ou seja, $2^{-p} \cdot 2^k = 2^{k-p}$, que é o total de experimentos a ser realizado e o valor de p corresponde ao número de geradores independentes (MONTGOMERY, 2001, p. 326). Por

exemplo, um experimento com seis fatores (A, B, C, D, E, F) e dois níveis para cada fator será representado por 2^6 , um total de 64 experimentos e se, por certa circunstância, esse experimento for reduzido a $(\frac{1}{2})^2$ do total, ou seja, $2^{-2} \cdot 2^6 = 2^{6-2}$, ele terá 16 experimentos.

3.3 DELINEAMENTO DE PRÉ-TESTE E PÓS-TESTE COM UM GRUPO

O Delineamento de Pré-Teste e Pós-Teste com um grupo é classificado como Pré-experimental. É um método de pesquisa utilizado para avaliar os efeitos de práticas de ensino com origem em inovação curricular ou para verificar os impactos de certo procedimento didático em determinado grupo de estudantes (COHEN; MANION; MORRISON, 2007). Pode ser representado, em notação, como:

Experimentação $O_1 \quad X \quad O_2$, em que:

1) O_1 corresponde aos resultados obtidos através da aplicação do Pré-Teste, o qual tem por objetivo medir, inicialmente, determinada variável específica dos sujeitos da pesquisa, antes de submetê-los a algum tratamento; 2) X é o próprio tratamento ministrado aos sujeitos, com o propósito de modificar o estado atual de O_1 ; 3) O_2 diz respeito aos resultados provenientes da aplicação do Pós-Teste, após submissão dos sujeitos a X . O que se pretende analisar é a diferença entre O_2 e O_1 , buscando atribuir sua causa ao tratamento.

Apesar de ser um método amplamente utilizado em pesquisa educacional, esse delineamento apresenta certa vulnerabilidade. Segundo Campbell e Stanley (1963), existem fatores internos e externos que podem ameaçar a validade desse procedimento: *eventos históricos* são passíveis de ocorrer durante o tratamento, como por exemplo, uma guerra, crise econômica, ou mudanças sociais que influenciem os sujeitos quando em tratamento; também há a questão do *amadurecimento* biológico e psicológico dos alunos com o passar do tempo. Inclusive, o próprio efeito do Pré-Teste pode ser uma causa de interferência, pois os estudantes, de alguma forma, podem ter sido influenciados pelos conceitos presentes no teste, ou até mesmo, terem aprendido durante sua realização, assim como podem ter relembrado conceitos assimilados no passado. Além de descreverem outras possíveis ameaças a esse delineamento de pesquisa, as quais denominam *hipóteses rivais*, os autores enfatizam que essa prática de pesquisa deve ser utilizada apenas quando não houver a possibilidade de se empregar outro tipo de delineamento. Podem ser citados outros meios como: o delineamento com grupo de controle, estudo de caso, delineamento

com grupo não equivalente, pesquisa-ação, etc (COHEN; MANION; MORRISON, 2007; PATHAK, 2008; CRESWELL, 2012).

3.4 VALIDAÇÃO DE INSTRUMENTOS

A validade é um antecedente fundamental do teste, assim como sua fidedignidade. Somente após o cumprimento dessas condições pode-se dizer que o instrumento está apto para avaliar (VIANNA, 1978).

Raymundo (2009) afirma que a validade de um teste começa a partir do instante em que se conjectura construí-lo até a fase de interpretação de seus resultados. Define-se validação como sendo o processo pelo qual se analisa a exatidão de uma predição ou dedução gerada a partir dos resultados de um teste. Valida-se, portanto, não exclusivamente o instrumento, mas a leitura que se faz das informações obtidas por meio de um processo característico. A autora se refere a três aspectos da validação de um teste:

a) Validade de conteúdo: trata-se do julgamento exercido sobre o instrumento para verificar se realmente o teste envolve conceitos específicos do que se pretende mensurar, ou se ele avalia assuntos pertencentes a outros conteúdos. Submete-se o instrumento a um júri especialista, que verifica a pertinência dos itens (questões) a seus respectivos conteúdos, a linguagem escrita, etc. Hernandez-Nieto (2001) propõe um Coeficiente de Validade de Conteúdo (CVC) para cada item do teste (CVC_i) e para o teste como um todo (CVC_t). São considerados, para o cálculo desse coeficiente, critérios como clareza de linguagem e pertinência prática. Para determinar o valor desses coeficientes são empregadas as seguintes equações:

1) A média das notas de cada item:

$$M_x = \frac{\sum_{i=1}^J x_i}{J} \quad (1)$$

Em que $\sum x_i$ corresponde à soma das notas dos juízes nos critérios e J a quantidade e juízes que avaliaram o teste.

- 2) Para o cálculo do coeficiente de validade de cada item adota-se:

$$CVC_i = \frac{M_x}{V_{máximo}} \quad (2)$$

Em que $V_{máximo}$ é a maior nota que um item poderia receber.

- 3) Para o cálculo do desvio viés do julgamento dos juízes sobre cada item (Pe_i), aplica-se:

$$Pe_i = \left(\frac{1}{J}\right)^J \quad (3)$$

- 4) A partir desse valor, pode-se determinar o valor do CVC final de cada item conforme a equação:

$$CVC_{t_i} = CVC_i - Pe_i \quad (4)$$

- 5) Por fim, para se determinar o valor do CVC total do teste calcula-se:

$$CVC_t = M_{CVC_{t_i}} - M_{Pe_i} \quad (5)$$

Em que $M_{CVC_{t_i}}$ é a média dos coeficientes de validade final de cada item e M_{Pe_i} a média dos vieses dos juízes nos itens do questionário.

O valor mínimo adotado e recomendado para os resultados de CVC_{t_i} e CVC_t é $\geq 0,70$ (CASSEPP-BORGES, V.; BALBINOTTI, M.A.A.; TEODORO, M.L.M., 2010).

b) Validade de critério: refere-se à capacidade que um instrumento de medida tem de prever o desempenho de um sujeito. Equivale a dizer que o desempenho do sujeito em outras instâncias passa a ser o critério de comparação com a medida obtida no teste. Há dois tipos de validade de critério, a *preditiva* e a *concorrente*. A diferença entre elas está no tempo de comparação do critério com os resultados do teste. Quando a comparação é praticamente simultânea, diz-se que a validade é do tipo concorrente; caso seja depois da obtenção das informações sobre o teste, é denominada preditiva. Há diferentes critérios que podem ser adotados para fazer frente aos resultados do teste: desempenho acadêmico; desempenho em treinamento especializado; desempenho profissional; diagnóstico psiquiátrico, etc (PASQUALI, 2003). Pode-se calcular a validade de critério estatisticamente. Quando a correlação entre os escores do teste (X) e os escores

do critério (Y) for alta, valida-se o teste para sua utilização (RAYMUNDO, 2009). Um método estatístico utilizado para verificar o grau de correlação entre X e Y é o coeficiente de correlação produto-momento, mais conhecido como coeficiente de correlação de Pearson ou também r de Pearson (RODGERS; NICEWANDER, 1988):

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2\right]^{1/2}} \quad (6)$$

A equação descreve r como a soma centralizada e padronizada do produto cruzado das duas variáveis (X e Y), onde $-1 < r < +1$. Quando o coeficiente tender a 1, diz-se que a correlação está próxima a uma correlação positiva perfeita, mais perto de -1, correlação negativa perfeita, zero (0) correlação nula. Para valores negativos, X e Y se relacionam de forma inversamente proporcional, quando positivos de forma diretamente proporcional.

Hinkle, Wiersma e Jurs (1998, p. 120) apresentam uma relação para a interpretação do coeficiente de correlação, dado seu valor, conforme a Tabela 2:

Tabela 2 – Regra para interpretação do coeficiente de correlação.

Valor da correlação	Interpretação
0,90 até 1,00 (-0,90 até -1,00)	Correlação muito alta
0,70 até 0,90 (-0,70 até -0,90)	Correlação alta
0,50 até 0,70 (-0,50 até -0,70)	Correlação moderada
0,30 até 0,50 (-0,30 até -0,50)	Correlação baixa
0,00 até 0,30 (-0,00 até -0,30)	Pouquíssima correlação

Fonte: Adaptado de Hinkle, Wiersma e Jurs (1998, p. 120).

É importante ressaltar que, um coeficiente de correlação de valor baixo pode ser tão significativo quanto um de correlação alta, pois a maioria das relações entre quaisquer variáveis estão distantes da perfeição. Knoke e Mee (2002) reiteram essa afirmação a respeito do coeficiente de correlação em pesquisa social. Segundo os autores, raramente uma variável independente explica mais do que 25 % a 30 % de outra variável independente. A maioria dos casos para menos de 2 % até 5 %. Para saber se a correlação entre as variáveis é forte, utiliza-se o r^2 , obtido através de r . Os autores classificam essa relação (Tabela 3):

Tabela 3 – Regra de interpretação do coeficiente de correlação.

r	r^2	Frequência em pesquisa social
0,50	0,25	Raramente
0,55	0,30	Raramente
0,14	0,02	Maioria dos casos
0,22	0,05	Maioria dos casos

Fonte: (KNOKE; MEE, 2002).

Sendo que r^2 corresponde à variação total comum, em porcentagem, aos elementos parte dos pares analisados. Por exemplo, supondo que a correlação entre os dados estimados por um modelo qualquer e os dados obtidos em observação seja $r = 0,50$, isso quer dizer que o modelo consegue explicar 25 % dos dados observados. Quanto maior o valor de r^2 , mais o modelo se ajusta aos dados observados.

De acordo com Rosner (1995, p. 507), a significância do coeficiente de correlação de Pearson é uma função do coeficiente de correlação de Pearson e o tamanho da amostra. O autor propõe um meio para analisar a significância do coeficiente de correlação em relação ao tamanho da amostra, ou seja, testa a hipótese nula (H_0) de que o coeficiente de correlação da população é diferente de zero:

$$H_0: \rho = 0 \quad \text{VS} \quad H_1: \rho \neq 0,$$

Em que, ρ é o coeficiente de correlação da população.

Considerando o teste $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, com r sendo o coeficiente de correlação e dada uma distribuição bicaudal com certo nível de α , rejeita-se H_0 se:

$$t > t_{n-2, 1-\alpha/2} \quad \text{ou} \quad t < -t_{n-2, 1-\alpha/2}$$

Assim, para uma distribuição bicaudal com $\alpha = 0,05$, a significância estatística de r ($p < 0,05$) é atingida quando:

$$|t| > t_{n-2, 1-0,05/2} = |t| > t_{n-2, 0,975}$$

Logo, dado qualquer tamanho da amostra (n) e conhecido o valor de r , determina-se a significância do coeficiente caso:

$$\left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| > t_{n-2, 1-\alpha/2}$$

ou

$$r > \frac{t_{n-2, 0,975}}{\sqrt{n-2 + (t_{n-2, 0,975})^2}} \quad (7)$$

c) **Validade de construto:** permite determinar qual aspecto educacional explica a variância do teste, seu significado, grau de homogeneidade, etc. A validade de construto não se limita apenas à validação do teste, seu alcance é muito maior, ou seja, o objetivo principal é validar a teoria em que se apoiou a construção do instrumento (KERLINGER; LEE, 1973). Para Cronbach e Meehl (1955), esse tipo de validade é empregado somente quando um teste será interpretado como a medida de algum atributo ou qualidade, a qual não está muito bem definida, onde há incerteza sobre outros construtos que podem estar interferindo nos resultados do teste. Neste caso, métodos estatísticos podem ser aplicados. Vianna (1983) afirma que a Análise Fatorial Exploratória é um método promissor, que deve ser explorado na validação de construtos.

d) **A fidedignidade** de um teste diz respeito à solidez de seus resultados. Para um teste aplicado múltiplas vezes a um mesmo grupo, espera-se o mesmo resultado. Quando se aplica o mesmo teste várias vezes, sob a mesma condição, aos mesmos sujeitos, e eles obtêm pontuações muito diferentes, então não há consistência nos resultados e o teste não é confiável. Pode-se calcular a fidedignidade de um teste através do método da metade. Divide-se o teste em dois, como se fossem dois similares. Estima-se a fidedignidade pela fórmula de Spearman-Brown (VIANNA, 1978, p. 153):

$$r_{xx} = \frac{\frac{\sum XY}{N} - \bar{X} \bar{Y}}{S_X S_Y} \quad (8)$$

Em que r_{xx} corresponde à fidedignidade do teste com metade do número de itens;

$\sum XY$ é o produto entre cada metade dos escores;

N é o número de sujeitos participantes do teste;

$\bar{X} \bar{Y}$ é o produto entre as médias das metades dos escores;

e $S_X S_Y$ é o produto entre os desvios padrões das metades dos escores.

Aplicando a fórmula da profecia de Spearman-Brown para todo o teste, ou seja, multiplicado por 2 obtém-se:

$$r_{(n)x} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}} \rightarrow r_{(2)x} = \frac{2r_{xx}}{1 + (2 - 1)r_{xx}} \quad (9)$$

Em que $r_{(2)x}$ é a fidedignidade para todo o teste;

r_{xx} corresponde à fidedignidade do teste com metade do número de itens;

e n é o número de vezes que o teste deve ser aumentado;

O coeficiente de fidedignidade varia de 0 a 1, quanto mais próximo seu valor estiver de 1, mais confiável é o teste. Pasquali (2003) afirma que um coeficiente de fidedignidade próximo a 0,80 é considerado razoável e abaixo de 0,70 é insuficiente para determinar a fidedignidade de um teste.

3.5 ANÁLISE FATORIAL EXPLORATÓRIA

A maior parte da literatura divide a Análise Fatorial em dois tipos, exploratória e confirmatória. A Análise Fatorial Exploratória (AFE), geralmente, é aplicada em um estágio preliminar da pesquisa, buscando compreender melhor os dados observados. Nela, analisa-se a relação entre determinadas variáveis, verificando padrões de correlação. Na Análise Fatorial Confirmatória (AFC), já se faz uma ideia sobre a relação entre os fatores e as variáveis. Realiza-se, então, um teste de hipótese com o objetivo de confirmar as relações previamente especuladas (FIGUEIREDO FILHO; SILVA JÚNIOR, 2010; TABACHNICK; FIDELL, 2013).

A AFE é um procedimento amplamente utilizado no desenvolvimento e avaliação de instrumentos psicológicos de medida (FLOYD; WIDAMAN, 1995; FABRIGAR et al., 1999). Sua principal intenção é determinar a quantidade e a espécie das variáveis latentes (fatores) em certo conjunto de dados observados (BROWN, 2015).

Willians, Brown e Osman (2010) provêm um roteiro de cinco passos, com base empírica, para a realização da AFE. As etapas se resumem nas seguintes sentenças: (1) Os dados são passíveis de análise fatorial? (2) Como os fatores serão extraídos? (3) Que

critério auxiliará na determinação da extração dos fatores? (4) Seleção e aplicação do método de rotação; (5) Interpretação e rotulação dos fatores.

No primeiro passo, deve-se ponderar o tamanho da amostra. Uma amostra com 100 sujeitos é considerada pobre, 200 razoável, 300 boa, 500 muito boa e 1000 ou mais excelente (COMREY, 1973). Outra parte dessa etapa é a verificação da fatorabilidade da matriz de correlação.

A matriz de correlação diz respeito, na AFE, às relações entre as variáveis individuais. Coeficientes de correlação com cargas de $\pm 0,30$ são mínimas, $\pm 0,40$ são importantes e $\pm 0,50$ praticamente significantes; caso não haja correlações acima de 0,30 na matriz de correlação, é prudente avaliar se a análise fatorial é o método estatístico correto a se utilizar (HAIR et al., 2009). Dois métodos muito utilizados para verificação da suscetibilidade dos dados à análise fatorial são: Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett (KAISER, 1960, 1970; BARTLETT, 1950).

O critério de KMO (KAISER, 1970, p. 405) verifica se os dados observados são compatíveis ao modelo de análise fatorial empregado. É representado pela equação:

$$KMO = \frac{\sum_{j \neq k} \sum r_{jk}^2}{\sum_{j \neq k} \sum r_{jk}^2 + \sum_{j \neq k} \sum q_{jk}^2} \quad (10)$$

Em que,

r_{jk}^2 representa a correlação simples entre as variáveis externas à diagonal;

q_{jk}^2 é a correlação parcial entre as variáveis ao quadrado.

O valor de KMO varia de 0 a 1, sendo 0,50 o mínimo aceitável (HAIR et al., 2009). Kaiser (1974) sugere a seguinte escala para os valores de KMO: $\geq 0,9$ excepcional, $[0,8, 0,9)$ meritório, $[0,7, 0,8)$ mediano, $[0,6, 0,7)$ medíocre, $[0,5, 0,6)$ miserável e $< 0,5$ inaceitável.

O teste de esfericidade de Bartlett tem distribuição aproximada de Qui-Quadrado e verifica a hipótese nula de que não há correlação entre os dados, ou seja, de que a matriz

correlação é matriz identidade. Sendo a matriz identidade, diz-se que a aplicação do modelo fatorial é inadequada (BARTLETT, 1954, p. 297). Esta estatística é calculada por:

$$\chi^2 = -B \log_e |R|$$

em que,

$$B = n - 1 - \frac{(2p + 5)}{6}$$

logo,

$$\chi^2 = - \left[n - 1 - \frac{(2p + 5)}{6} \right] \ln |R| \quad (11)$$

com grau de liberdade: $\frac{p(p - 1)}{2}$

em que,

$|R|$ é a determinante da matriz de correlação;

n é o tamanho da amostra;

p corresponde ao número de variáveis.

Valores elevados para o teste de esfericidade de Bartlett indicam a rejeição da hipótese nula de que a matriz correlação é identidade. Para isso, considerando um nível de significância de 95 %, o valor da estatística p (p -value) deve ser menor do que 0,05 e χ^2 calculado deve ser maior que o valor crítico, definido na tabela de Qui-Quadrado (THOMPSON, 2004; TABACHNICK; FIDELL, 2013).

O segundo passo corresponde à forma como os fatores serão extraídos. Na AFE, os principais métodos empregados na realização dessa tarefa são: a Análise das Componentes Principais (ACP), Análise Fatorial tradicional (AF), Principais eixos fatoriais, Máxima Verossimilhança, Fatoração alpha, etc. Apesar de a ACP e a AF serem métodos parecidos e mais utilizados, produzindo resultados similares, existe diferença entre eles, mas ambos são aplicados para identificar grupos de variáveis observadas (THOMPSON, 2004; O'ROURKE et al., 2012; TABACHNICK; FIDELL, 2013).

Para execução do terceiro passo, há vários critérios definidos por diferentes autores que auxiliam na extração dos fatores: o critério de Kaiser, que exclui autovalores

maiores do que 1 e mantem menores do que 1 (KAISER, 1960); o *Screeplot* (CATTELL, 1966); a porcentagem de variância total acumulada; etc.

A variância total acumulada corresponde à soma das variâncias de todas as variáveis observadas (O'ROURKE et al., 2012, p. 14). Apesar de, em ciências naturais, a porcentagem de variância acumulada ter seu valor aceitável em 95 %, quando se trata de ciências sociais, o valor mínimo considerado satisfatório pode estar entre 50 % e 60 % (PETT; LACKEY; SULLIVAN, 2003; HAIR et al., 2009). O'Rourke (2012) descreve outro parâmetro aplicado nas pesquisas: individualmente, autovalores considerados devem apresentar pelo menos 5 % a 10 % de variância acumulada e, coletivamente, 70 % a 80 %.

O *Screeplot* é um gráfico subjetivo às considerações do pesquisador e seu uso tem sido discutido (KAISER, 1970; FABRIGAR et al., 1999). Sua interpretação é simples: **1)** Observar a reta y , que será o parâmetro de inclusão ou exclusão das Componentes Principais; **2)** Atentar para o fato de que o número de componentes a serem retidos está acima desta linha (CATTELL, 1966). O Gráfico 2 é um exemplo e neste caso seriam retidas apenas a primeira e a segunda componente, pois estão acima da reta.

Gráfico 2 – *Screeplot*.

Fonte: Próprio do autor.

No quarto passo, determina-se o método de rotação, empregado para redistribuir, de forma mais adequada, cargas fatoriais sobre seus respectivos fatores, tornando a interpretação mais simples e confiável. Quando as variáveis estão relacionadas a um único fator, diz-se que o conjunto de dados é *unidimensional*. A finalidade da rotação é maximizar itens com altas cargas fatoriais e minimizar itens com cargas fatoriais baixas; os tipos de rotação comumente empregados são a ortogonal e oblíqua. A técnica ortogonal

mais utilizada em análise fatorial é a de Varimax, porém existem outras do mesmo tipo: Quartimax, Equimax, Varimin, etc. Algumas das rotações oblíquas são: Promax, Oblimin, etc. Escolhido e aplicado o método de rotação, o pesquisador então examina os itens que não têm carga suficiente ou que não estão aptos a serem designados a um fator; estes itens devem então ser descartados (WILLIAMS; BROWN; OSMAN, 2010).

A fase final, da interpretação e rotulação, é onde o pesquisador examina quais variáveis são atribuídas ao fator, dando ao fator nome. Por exemplo, um fator pode ter suas maiores cargas sobre quatro variáveis que se relacionam com o estado de depressão; por isso, o pesquisador irá rotular essa variável como “estado de depressão”. Em geral, quando pelo menos duas ou três variáveis têm maior carga sobre um único fator, pode-se atribuir a elas uma interpretação significativa (HENSON, 2006; ISAAC; MICHAEL, 1995).

Através da AFE, também pode se verificar a unidimensionalidade de instrumentos psicológicos de medida, que é pressuposto no emprego da Teoria Clássica dos Testes ou da Teoria de Resposta ao Item (PASQUALI, 2003). Para atingir a unidimensionalidade, tanto na TCT (Teoria Clássica dos Testes) como na TRI (Teoria de Resposta ao Item), supõe-se que os itens do teste estejam medindo um único construto. Como a TCT baseia-se no escore total, acredita-se que os itens tratem do mesmo conceito, caso contrário não seria possível a soma das pontuações nos itens, conseqüentemente não se obteria o escore total. Na TRI, a base é o traço latente, portanto os itens devem relacionar-se ao mesmo traço latente. Métodos comumente aplicados na verificação da unidimensionalidade são a AF e ACP, que utilizam a matriz de correlação dos dados, buscando verificar se há possibilidade de redução dos itens do teste a uma única dimensão, denominada fator; isso constatado, afirma-se que os itens são unidimensionais.

A relação do item com o fator é obtida através da correlação ou covariância, chamada de carga fatorial. Itens com carga fatorial alta caracterizam-se unidimensionais, com carga baixa devem ser excluídos. Cargas fatoriais altas apresentam um valor de pelo menos 0,30 (PASQUALI, 2003, p. 116). Outro parâmetro sugerido por um estudo da fundação Monte Carlo (GUADAGNOLI; VELICER, 1988) é: **1)** Componentes/fatores com quatro ou mais cargas acima de 0,60 (em valor absoluto) são significantes, independente do tamanho da amostra; **2)** Componentes/fatores com cerca de 10 ou mais cargas abaixo de 0,40 serão significantes apenas se o tamanho da amostra for por volta de

150 ou mais; **3)** Componentes/fatores com apenas poucas cargas próximas a 0,40 não devem ser interpretadas, ao menos que a amostra seja de pelo menos 300.

Outro critério que auxilia na verificação da unidimensionalidade é o da comunalidade dos itens. Ele corresponde à porcentagem de variância em uma variável observada. Variáveis com cargas altas apresentam uma comunalidade alta (O'ROURKE et al., 2012). De acordo com Velicer e Fava (1998), para que uma comunalidade seja alta, deve apresentar um valor de 0,80 ou mais, o que é muito difícil de ocorrer em dados reais. Em ciências sociais, os valores geralmente encontrados estão entre baixo (0,40) e moderado (0,70) (OSBORNE; COSTELLO, 2009).

3.6 TEORIA CLÁSSICA DOS TESTES

A Teoria Clássica dos Testes (TCT) tem uma longa história de serviço na tecnologia de construção de testes (NOVICK, 1965). Baseada na concepção de Spearman, que o resultado do teste é dado pelo escore real e o erro, a TCT foi introduzida no início do século XX (HAMBLETON; LINDEN, 1982). A maior parcela dos psicólogos, cientistas sociais e comportamentais, admite que a obra de Gulliksen (1950), *Theory of Mental Tests*, apresenta a TCT em seu estado mais detalhado. Gulliksen modela a teoria conforme a equação matemática $X = T + E$. Onde X é o escore observado de um sujeito, T o escore real desse sujeito e E os erros de medida ocorridos durante o processo. No diagrama da Figura 2 está relação entre as variáveis do modelo.

Figura 2 – Componentes do escore real.

Fonte: Adaptado de Pasquali (2003, p. 70).

Por meio da união do escore total (T) com os erros (E) é obtido o escore observado (X).

A TCT busca explicar o resultado final de um teste, ou seja, a soma das respostas às questões do instrumento avaliativo, denominado escore total (T). Por exemplo, o T de um teste com 20 questões corresponderia à soma dos itens corretamente respondidos. Um sujeito que acertasse 15 das vinte questões obteria um $T=15$. A TCT se deteria em esclarecer qual é o significado do $T=15$ para o sujeito. Outra função relevante da TCT é, através da Estatística, conter a amplitude dos erros derivados do processo de medida (PASQUALI, 2009). Esses podem ter origem em diferentes fontes: irregularidades no teste, fatos históricos, características dos sujeitos, etc. (CAMPBELL; STANLEY, 1963).

Segundo Hambleton e Jones (2005), muitos trabalhos sobre a TCT têm se concentrado em modelos fundamentados no escore do teste, ou seja, relacionam o escore real ao total e não o escore dos itens ao escore real (contrastando com a Teoria de Resposta ao Item – TRI). Para eles, as estatísticas desenvolvidas nesses modelos (dificuldade do item, grau de discriminação do item, média dos escores, correlação, confiabilidade, desvio padrão) dependem dos sujeitos em exame e essa dependência reduz a utilidade dos resultados. Ressaltam ainda que os resultados são mais úteis quando os examinandos pertencem à mesma população para qual o teste foi desenvolvido; que a TCT é incapaz de fazer previsões baseadas no desempenho do estudante a partir dos itens do teste, fornecendo apenas previsão apoiada no resultado total do teste (PASQUALI, 2003). Isso difere a TCT da TRI, que gera previsão fundamentada no desempenho dos sujeitos nos itens do teste, com base nas habilidades requeridas (traços latentes) pelos itens do teste (STEYER, 2001).

De acordo com Pasquali (2003), a dificuldade dos itens do teste é dada pela porcentagem de acertos em cada item (p_i); e o índice de discriminação de um item (D) é representado pela diferença entre a proporção de acerto do grupo de indivíduos com notas (escores) altas (p_i^A) e a quantidade de acertos do grupo com notas mais baixas (p_i^B):

$$D = p_i^A - p_i^B. \quad (12)$$

A maioria dos pesquisadores divide os dois grupos da seguinte forma: 0 % até 27 % de acertos é o Grupo Inferior (GI) e de 73 % a 100 % o Grupo Superior (GS), isso em uma curva de distribuição normal (KRISHNAN, 2013; KELLEY, 1939).

Ebel e Frisbie (1991) sugerem a seguinte tabela para os índices de discriminação:

Tabela 4 – Interpretação do índice de discriminação de um item.

Discriminação (D)	Interpretação do Item
$D < 0,19$	Fraco e precisa de revisão;
$0,20 < D < 0,29$	À margem do desejado, e precisa de alguma revisão;
$0,30 < D < 0,39$	Razoavelmente bom, mas possivelmente sujeito à melhoria;
$D > 0,40$	Muito bom.

Fonte: Ebel e Frisbie (1991).

O procedimento da divisão dos alunos em grupos de desempenho é conhecido como “Regra 27” de Kelley, como mostrado no Gráfico 3.

Gráfico 3 – Representação da “Regra 27”.

Fonte: Próprio do autor.

Onde $n1$ representa a maior nota entre os 27 % piores alunos e $n2$ a menor nota entre os 27 % melhores alunos.

Outra forma de se obter o índice de discriminação de um item é pelo Teste t:

$$t = \frac{\bar{X}_s - \bar{X}_i}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_i^2}{n_i}}} \quad (13)$$

Em que:

$\bar{X}_s$ e $\bar{X}_i$ são as médias do grupo superior e inferior;

s_s^2 e s_i^2 são as variâncias do grupo superior e inferior;

n_s e n_i é o número de sujeitos nos dois grupos.

Com graus de liberdade $gl = n_s + n_i - 2$ (PASQUALI, 2003, p. 134).

Esse teste compreende a análise da diferença entre as médias obtidas pelo grupo superior e inferior, necessitando das respectivas médias e variâncias. O teste consiste em rejeitar a hipótese nula (H_0) dado certo nível de significância (α). Nesse caso H_0 afirma que não há diferença entre as médias do Grupo Superior e Inferior. H_0 será rejeitada quando o valor de t for maior que α .

Quando as variáveis (questões) no teste são dicotômicas, isto é, quando os itens são do tipo certo (1) ou errado (0), pode-se utilizar a Correlação Ponto-bisserial, para cálculo do índice de discriminação, conforme a equação:

$$r_{pb} = \frac{\bar{X}_A - \bar{X}_t}{s_T} \sqrt{\frac{p}{q}} \quad (14)$$

Em que:

$\bar{X}_A$ é a média dos sujeitos que acertaram o item;

$\bar{X}_t$ é a média total do teste;

s_T é o desvio padrão do teste;

p é a proporção de sujeitos que acertaram o teste;

$q = 1 - p$ (PASQUALI, 2003, p. 135).

A interpretação da correlação ponto-bisserial é similar a do coeficiente de correlação de Pearson, descrita anteriormente.

Outro critério aplicado para verificar a dificuldade do teste sobre determinada população é o de assimetria, ou seja, o desvio ou afastamento da simetria (TOLEDO; OVALLE, 1983). Na Figura 3 estão os possíveis casos para uma distribuição normal.

Figura 3 – Assimetrias e Simetria.

Fonte: Adaptado de Toledo (1983, p. 249 – 250).

Para cada um dos casos, há um enquadramento em relação ao valor da média ($\bar{X}$), da moda (M_o) e mediana (M_d): a) para $M_o < M_d < \bar{X}$ ocorre assimetria positiva; b) $\bar{X} < M_d < M_o$ assimetria negativa; c) $\bar{X} = M_d = M_o$ simetria (SPIEGEL, 1978, p. 121; TOLEDO; OVALLE, 1983, p. 249). O emprego desse conceito auxilia, na TCT, a caracterizar o teste e verificar sua adequação à população para o qual foi construído.

Uma opção adicional para se determinar a simetria de uma distribuição normal é o coeficiente de assimetria de Karl Pearson (TOLEDO; OVALLE; 1983):

$$\alpha_3 = \frac{E[(X - \mu)^3]}{\sigma} \quad (15)$$

Em que, E é o valor esperado da variável X ; μ é a média de X ; e σ é o desvio padrão de X . Para $\alpha_3 < 0$, $\alpha_3 > 0$ e $\alpha_3 = 0$, tem-se respectivamente, assimetria negativa, positiva e simetria.

Além da aplicação dos conceitos da TCT discutidos anteriormente, empregou-se também, nesta pesquisa, a TRI, abordada no próximo item.

3.7 TEORIA DE RESPOSTA AO ITEM

De acordo com os relatos de Bock (2005), a Teoria de Resposta ao Item (TRI) provavelmente teve origem com os trabalhos de Thurstone (1925), Fisher e Yates (1963), Lazarsfeld (1950) e outros; porém, foi nos trabalhos de Lord (1952) que a teoria passou a ter maior formalidade. A TRI veio se estabelecer e substituir *maior parte* da teoria clássica (PASQUALI, 2003; SARTES; SOUZA-FORMIGONI, 2013). Um dos motivos que conduz essa tendência é que a TCT se baseia apenas no escore total do teste e depende dos sujeitos em que se aplica o teste, o que não ocorre na TRI, a qual se fundamenta nos traços latentes dos sujeitos. A TRI se funde em dois axiomas:

- a) O desempenho do examinando nas questões do teste é explicado conforme os traços latentes dos sujeitos (competências, habilidades, etc.).
- b) O Desempenho é o efeito e os traços latentes a causa.

A função que descreve a relação entre o desempenho e o traço latente é a monotônica crescente, denominada Curva Característica do Item (CCI), exibida no Gráfico 4 (BAKER, 2001; MACDONALD; PAUNONEN, 2002; PASQUALI, 2003).

Gráfico 4 – Exemplo de Curva Característica do Item (CCI).

Fonte: Próprio do autor.

Na curva característica, a probabilidade ($P\theta_i$) de o sujeito acertar o item está relacionada à habilidade requerida do item ao aluno (θ_i). Quanto maior a aptidão do aluno naquela habilidade, maior a probabilidade de acerto no item. A capacidade de discriminação do item é dada pela inclinação da curva (valores de **a**). A projeção de **a**

sobre o eixo das abscissas (**b**) corresponde à dificuldade desse item quando a probabilidade de acerto for de 50 % (SARTES; SOUZA-FORMIGONI, 2013).

O modelo matemático da CCI é a forma cumulativa da função logística. Foi derivada em 1844 e tem sido muito utilizada na biologia para modelar o crescimento de plantas e animais, do nascimento à maturidade. Foi empregada como curva característica do item depois de 1950 e, pela sua forma simples, tornou-se o modelo preferido (BAKER, 2001). A equação para o modelo de dois parâmetros logísticos é dada a seguir:

$$P\theta = \frac{1}{1 + e^{-L}} + \frac{1}{1 + e^{-a(\theta-b)}} \quad (16)$$

Onde,

a é o parâmetro de discriminação;

$L = a(\theta - b)$ é o desvio logístico;

θ o nível de habilidade. (BAKER, 2001)

O parâmetro de dificuldade b é definido no ponto da escala de habilidade onde a probabilidade de responder corretamente o item é 0,5 (50 %). Varia no intervalo de $-\infty \leq b \leq +\infty$. Na prática, seus valores compreendem $-3 \leq b \leq +3$, quanto maior o valor do índice, mais difícil é o item. Já o parâmetro de discriminação do item, pode ser definido como proporcional à inclinação da curva quando $\theta = b$, ou seja, a inclinação no momento de inflexão. O alcance teórico desse valor está entre $-\infty \leq a \leq +\infty$, mas o intervalo usual prático é de -2,80 a +2,80. A maior parte dos testes irá discriminar um item positivamente, ou seja, a probabilidade de se responder ao item corretamente aumenta à medida que o grau de habilidade se eleva, mas itens podem apresentar discriminação negativa. Isso quer dizer que a probabilidade de se acertar o item decresce conforme o nível de habilidade cresce de baixo para cima, isto é, sujeitos com maior aptidão têm menor chance de acerto, enquanto os de menor habilidade teriam maior probabilidade, o que soa estranho. Por isso, em geral, adotam-se somente os valores positivos (BAKER, 2001; PASQUALI; PRIMI, 2003).

A TRI é um instrumento poderoso para avaliar instrumentos psicológicos, mas necessita que seu modelo se adeque aos dados coletados. Além da unidimensionalidade, há

inúmeras ferramentas disponíveis para auxiliar nessa tarefa, por exemplo, o Teste de Qui-Quadrado (χ^2) e o *Scatterplot* (PASQUALI, 2003).

O Teste de Qui-Quadrado, de ajuste do modelo aos dados (*goodness-of-fit*), é determinado pela seguinte equação:

$$\chi^2 = \sum_{i=1}^k \frac{(P_{1i} - P_{0i})^2}{P_{0i}} \quad (17)$$

Onde, k é o número de células (sujeitos);

P_{1i} são os dados observados;

P_{0i} são os dados estimados;

Para ajustes do modelo ao teste, tem-se $gl = k - 1$ graus de liberdade;

Para tabelas de contingência, $gl = (a - 1)(b - 1)$ graus de liberdade, onde a e b são a quantidade de níveis nas duas variáveis, ou seja, a o número de linhas e b o número de colunas da matriz de dados (COHEN, 1992).

No teste de Qui-Quadrado, são assumidas duas hipóteses: a) Hipótese nula (H_0) de que há consistência do modelo com os dados observados; b) Hipótese alternativa (H_a) de que não há consistência do modelo com os dados observados. Quando o valor observado da estatística Qui-Quadrado excede o valor crítico (obtido através da tabela de Qui-Quadrado), a hipótese nula de que o modelo se ajusta aos dados é rejeitada e, portanto, deve-se buscar um modelo com melhor ajuste (HAMBLETON; SWAMINATHAN; ROGERS, 1991).

O *Scatterplot* é um gráfico de dispersão que correlaciona duas variáveis, no caso dessa pesquisa, os dados observados contra os dados estimados pelo modelo. Quanto maior for a linearidade, mais próximo do real está a estimacão; e a adequação do modelo aos dados é aceitável (PASQUALI, 2003; LORD, 1974).

Outro parâmetro utilizado para verificar a adequação do modelo aos dados e aos sujeitos da pesquisa é o chamado índice de ajuste, que consiste em estatísticas denominadas *infit* e *outfit*. O *infit* verifica as respostas inesperadas dos sujeitos nos itens, ou seja, se sujeitos com proficiência (nível de θ) maior que a dificuldade (b) do item estão

errando no item, quanto mais erros inesperados, maior o valor de *infit*. O *outfit* tem praticamente a mesma função, mas tem por base os erros extremos, ou seja, pessoas com capacidade muito alta errando itens fáceis e pessoas com capacidade muito baixa acertando itens difíceis. Seus valores esperados estão próximos de 1,0; acima de 1,5 podem ser considerados problemáticos (LINACRE, 2012, p. 320; HUTZ; BANDEIRA; TRENTINI, 2015).

4 METODOLOGIA

Neste trabalho foi utilizada a forma de pesquisa mista, a saber, um Pré-experimento (quantitativo) conjunto a observações (qualitativo). O Pré-experimento corresponde ao de Delineamento de Pré-Teste e Pós-Teste com um grupo, conforme abordado na revisão bibliográfica. Para o cumprimento do Pré-experimento, foi construído e validado um teste de Proficiência em Estatística e Probabilidade (TEPEP), a fim de ser utilizado como Pré-teste e Pós-teste, na obtenção de dados quantitativos. Para a análise qualitativa, foram feitos registros das atividades dos alunos ao longo da aplicação da sequência didática. Ao fim das atividades, os resultados quantitativos foram analisados em paralelo aos resultados qualitativos. Neste capítulo, serão apresentadas informações sobre os sujeitos da pesquisa e do cenário de aplicação; as etapas de validação do TEPEP; a forma de aplicação do projeto com sua sequência didática (materiais utilizados, recursos tecnológicos, forma de abordagem dos conteúdos); e a ferramenta utilizada para a análise dos dados e composição dos resultados.

4.1 PÚBLICO ALVO

Participaram desta pesquisa um total de 391 alunos do Ensino Fundamental II (EFII). Desse total, 374 eram de 9º ano, cursando escolas públicas e privadas da cidade de Lorena, Cruzeiro e Guaratinguetá, do Estado de São Paulo. A função desse grupo, de 374 alunos, foi responder ao TEPEP para que suas pontuações fossem utilizadas, exclusivamente, no cumprimento de todas as etapas de validação do teste, menos uma, a validade de critério (vide item 3.4, p. 13). O outro grupo, de dezessete (17) alunos, de 8º ano, com faixa etária entre 13 a 15 anos e membros de uma escola pública do Município de Lorena, foi sujeito na aplicação da metodologia, desenvolvida nesta dissertação. Além disso, durante a aplicação da metodologia, esses 17 alunos também responderam ao TEPEP, quando em Pré-Teste e Pós-teste. Suas pontuações, coletadas no Pós-teste, foram utilizadas para atingir a validade de critério do TEPEP. A escolha do 8º ano, como sujeito na aplicação da metodologia, deve-se a possibilidade de abordagem mais profunda dos conteúdos da Estatística e Probabilidade, pois geralmente, é nessa etapa que o estudo do assunto torna-se mais intenso.

4.2 TESTE DE PROFICIÊNCIA EM ESTATÍSTICA E PROBABILIDADE

A elaboração do Teste de Proficiência em Estatística e Probabilidade (TEPEP) (Apêndice A) buscou suprir a necessidade do delineamento de pesquisa aqui adotado, que requisita a medida de certa característica dos sujeitos antes e depois de submetê-los a um tratamento, conforme descrito no item 3.3.

Como todo instrumento de medida deve estar devidamente calibrado, a validação do TEPEP seguiu conforme os critérios de validade descritos na revisão bibliográfica: conteúdo, critério e construto; pressupondo a aplicação da Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI). O fluxograma na Figura 4 apresenta os passos realizados para a validação do teste.

Figura 4 – Fluxograma do processo de validação.

Fonte: Próprio do autor.

Visto que a validação de um instrumento tem início desde o momento em que se pensa na sua elaboração, o processo de validação do TEPEP começou na escolha dos itens.

Inicialmente, foi organizado um banco de questões com aproximadamente 60 itens, de 5º, 7º e 9º anos do Ensino Fundamental, escolhidas de questionários das provas do SAEB (Prova Brasil) e dos relatórios pedagógicos anuais do SARESP (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo). Algumas das questões foram construídas pelo autor deste trabalho e todas as que foram escolhidas pautaram-se nos conteúdos dos Parâmetros Curriculares Nacionais (PCN), especificamente do bloco

Tratamento da informação, que diz respeito aos conceitos da Estatística e Probabilidade. De todas as questões do banco, foram selecionadas 15, para contemplar os objetivos específicos do projeto, descritos no capítulo 2 desta dissertação.

Para atingir a validade de conteúdo, submeteu-se o teste a um júri especialista, constituído por um conjunto de cinco professores universitários experientes, do curso de Licenciatura em Matemática, do Centro Universitário Salesiano de São Paulo (UNISAL), situado também no município de Lorena.

Os juízes avaliaram os 15 itens do teste segundo a Clareza de Linguagem (CL), Pertinência Prática (PP) e Relevância Teórica (RT), de acordo com a seguinte escala *likert*:

- 1- Inapropriado;
- 2- Pouquíssimo apropriado;
- 3- Pouco apropriado;
- 4- Apropriado;
- 5- Muito apropriado.

Clareza de Linguagem: diz respeito à linguagem utilizada nos itens, considerando o perfil da população que utilizará a ferramenta (alunos de 8º ou 9º ano). Deve se responder à pergunta “A linguagem é suficientemente clara para que a população em questão compreenda esse item? Em qual escala?”.

Pertinência Prática: refere-se à capacidade que o item tem para avaliar o nível de proficiência de alunos em Estatística e Probabilidade. A pergunta a ser respondida é “O item analisado é pertinente para avaliar o nível de proficiência dos alunos em Estatística e Probabilidade? Em que nível?”.

Relevância Teórica: Compreende o grau de adequação do conteúdo da questão à teoria estudada. Deve-se indagar “O conteúdo dessa questão é significativo em relação à proficiência em Estatística e Probabilidade? Em que grau?”.

Com base nas notas dadas pelo júri nos três critérios em cada item, aplicou-se o cálculo do Coeficiente de Validade de Conteúdo (CVC), com o objetivo de analisar estatisticamente a qualidade do teste sobre a perspectiva dos experientes professores.

Após a avaliação do júri, o teste foi aplicado a 374 alunos de 9º ano, de escolas públicas e particulares do município de Lorena, Guaratinguetá e Cruzeiro, todos do Estado

de São Paulo. Da cidade de Cruzeiro, apenas uma escola participou, sendo ela particular; em Guaratinguetá, também uma única escola, pública e estadual; a maior parte das escolas em que o teste foi aplicado é de Lorena, sendo duas particulares, uma estadual e três municipais. São duas as razões para a aplicação do teste em escolas particulares: 1) aumentar a quantidade da amostra estatística; 2) construir um teste para ser aplicado tanto em escolas públicas quanto em particulares. A aplicação do teste foi realizada pelos professores dessas escolas: dois professores de Matemática, um de Biologia e um de Arte. Todos foram instruídos sobre a aplicação do teste. O objetivo da etapa foi a obtenção de dados para serem utilizados na validação do TEPEP.

Para determinar a validade de construto, empregou-se a Análise Fatorial Exploratória, seguindo cinco passos: 1) verificar se os dados são passíveis de análise fatorial; 2) Realizar a extração dos fatores; 3) Utilizar critérios para auxiliar na extração e redução dos fatores; 4) Determinar e aplicar o método de rotação; 5) Interpretação e rotulação dos fatores.

Como pressuposto, para o cumprimento do primeiro passo, fez-se uso do método Kaiser-Meyer-Olkin (KMO) e do teste de esfericidade de Bartlett, buscando verificar a suscetibilidade da matriz correlação dos dados coletados à análise fatorial; a segunda etapa, da extração dos fatores (componentes), foi realizada através da Análise das Componentes Principais (ACP); e para auxiliar na extração e redução desses fatores (terceira etapa), empregou-se a técnica do *Screeplot* e da porcentagem de variância acumulada; para rotação das cargas fatoriais, aplicou-se a rotação ortogonal de Varimin; por fim, na última fase, de interpretação e rotulação dos fatores, verificou-se, com base nos resultados da ACP, a unidimensionalidade do teste, através do cálculo das cargas fatoriais dos itens e suas posições sobre seus respectivos fatores, assim como da comunalidade dos itens. Todos os cálculos foram realizados utilizando, como ferramenta tecnológica, o *software* estatístico *R*, abordado posteriormente neste documento.

A validade de critério foi analisada comparando o desempenho acadêmico dos 17 alunos com suas pontuações no TEPEP (Pós-Teste). O desempenho acadêmico foi calculado a partir das notas dos alunos em disciplinas cursadas no primeiro bimestre de 2016 (História, Geografia, Língua Portuguesa, Matemática, Ciências e Educação Física). Através desses dados, calculou-se o coeficiente de correlação de Pearson, para verificar a capacidade que o teste tem de discriminar alunos melhores dos piores.

Verificados os critérios de validade e com base nos resultados dos 374 alunos, aplicou-se a TCT e TRI, para caracterizar o TEPEP e, com base nisso, analisar o desempenho dos 17 alunos participantes da sequência didática desenvolvida.

4.3 A ESTATÍSTICA DOS TEMPOS DE VOOS DOS AVIÕES DE PAPEL

A concepção do projeto (A Estatística dos tempos de voos dos Aviões de Papel) teve inspiração no experimento de George Box (1992), que utilizou helicópteros de papel para ensinar Estatística avançada a alunos de engenharia. Em seu projeto, os alunos discutiram e determinaram as alterações que fariam nos helicópteros para permanecerem mais tempo no ar. Após terem discutido as alterações, aplicaram técnicas estatísticas de Planejamento de Experimentos para descobrir quais alterações aumentariam o tempo de voo dos helicópteros.

No projeto aqui descrito, a ideia foi similar, com duas diferenças: 1) a otimização foi feita em um modelo de avião de papel; 2) são abordados conceitos básicos de Estatística e Probabilidade.

Conforme as fundamentações próprias do Planejamento de Experimentos (PE) e da Aprendizagem ativa (AA), relativa ao projeto dirigido já discutido na revisão bibliográfica, desenvolveu-se uma Sequência didática (SD) para a aplicação do projeto (Quadro 1).

Quadro 1 – Bases metodológicas da sequência didática do Projeto.

AA	PE	SD
1. Reconhecer um problema e estabelecer uma meta.	1. Reconhecimento do problema	1. Reconhecer o problema; escolha do modelo de avião.
2. Reunir informações relevantes.	2. Escolha dos fatores, níveis e alcance.	2. Realização do <i>Brainstorm</i> .
3. Gerar possíveis soluções para o problema.	3. Seleção da variável resposta	3. Cálculo do total de experimentos; escolha do plano de experimento.
4. Listar restrições ao que pode ser feito.	4. Escolha do plano do experimento	4. Aula-treino.
5. Escolher uma solução possível.	5. Desenvolvimento do experimento	5. Sorteio das alterações.
6. Considerar possíveis fatores para uma solução detalhada.	6. Análise estatística.	6. Confecção dos aviões.
7. Criar uma solução detalhada.	7. Conclusão.	7. Realização do experimento.
8. Avaliar a solução final.		8. Análise estatística dos dados.
9. Recomendar um curso de ação.		9. Conclusão e recomendação

Fonte: Próprio do autor.

Nota-se que há semelhança entre os procedimentos da Aprendizagem ativa e Planejamento de Experimentos. Isso serviu de referência para a elaboração da sequência didática. O primeiro passo das três séries é exatamente o mesmo, ou seja, deve haver um problema a ser resolvido. As etapas subsequentes de AA (2, 3 e 4) se assemelham às etapas 2 e 3 de PE e 2 de SD, pois dizem respeito às primeiras considerações realizadas sobre o experimento. Percebe-se que, em 5 e 6 de AA, há similaridade com 4 de PE e 3 e 5 de SD, onde se determinam os fatores a serem analisados, assim como o plano de solução; em AA, não se observa uma etapa literalmente prática, de campo, o que acontece em 5 de PE e 4, 6 e 7 de SD. Porém, na criação da solução detalhada de AA, aparecem etapas parecidas em 6 de PE e 8 de SD, que tratam da análise estatística dos dados do experimento. Por fim, 8 e 9 de AA, estão diretamente relacionadas à conclusão e recomendação de um curso de ação, que ocorre em 7 de PE e 9 de SD. Independente do experimento que se pretenda realizar, as etapas da sequência didática permanecem estáticas, não sendo obrigatória a realização do passo 4, mas recomenda-se, pois os alunos conseguem observar e verificar o que podem ou não fazer para que os resultados obtidos no experimento sejam mais confiáveis.

Para o planejamento das etapas da sequência didática, utilizou-se como base a Aprendizagem colaborativa. A seguir, são descritas cada uma das etapas:

1. Reconhecer o problema: essa fase foi o momento de introdução do projeto. Quatro aviões (Anexos de A a D), confeccionados de forma caprichosa, foram postos em um local visível a todos os alunos. Os modelos foram: o D'artgnan, Glider, Arumakan (LEE, K.; MITCHELL, 2013) e Dart (FOLD'NFLY, 2015). Posteriormente, compartilhou-se com os alunos o seguinte problema *“Quais alterações podem ser feitas em um modelo de avião de papel para que ele permaneça mais tempo voando?”*. A turma sugeriu suas ideias iniciais, mas foram formalmente aceitas na etapa 2. Permitiu-se que os alunos escolhessem, por votação, apenas um dos modelos exibidos. Eles tiveram contato, um por um, com os quatro aviões, para escolherem de forma mais coerente. Os alunos dessa classe decidiram pelo modelo Arumakan (Anexo A). Vale observar que os modelos foram selecionados e confeccionados com antecedência pelo professor, para ter uma ideia do tempo e do grau de dificuldade da construção dos modelos. A seleção seguiu um critério: os modelos não podiam ser difíceis de confeccionar. Modelos complexos são demorados de se aprender e fazer, o que demanda tempo. A etapa de reconhecimento do problema durou, no máximo, 1h/aula (50min).

2. *Brainstorm*: é a etapa onde os alunos determinaram as alterações a serem feitas no modelo escolhido. Foi estudado pelo docente, com antecedência, como se realiza a sessão de *Brainstorm* de forma correta. Foi elaborada uma apresentação no *software* PowerPoint para orientar os alunos sobre as fases da sessão e como proceder em cada uma. No início, os alunos compartilharam suas ideias, mas seus colegas não puderam criticá-las; elas foram anotadas em um lugar visível a todo o grupo de alunos. Podiam ser feitas perguntas com a finalidade de esclarecer ou definir as ideias dadas. Após se esgotarem todas as sugestões sobre as alterações, os alunos escolheram as mais interessantes; a escolha foi realizada pela sala como um único grupo. Optaram por: dobrar a asa do avião, reduzir o tamanho do avião, usar papel sulfite e novo-antílope e deixar o corpo do modelo fixo; como no Quadro 2.

Quadro 2 – Alterações sugeridas pelos alunos.

Alterações	Nível 1 (-)	Nível 2 (+)
Corpo fixo	Não	Sim
Asa dobrada	Não	Sim
Tamanho do corpo	Original	Reduzido
Tipo de papel	Sulfite	Novo antílope

Fonte: Próprio do autor.

Os sinais dos níveis (– e +) correspondem às configurações originais do avião e às alterações a realizar, respectivamente. Essa forma de organização do quadro é própria do Planejamento de experimentos. O tempo necessário para realizar todo o *Brainstorm* foi de 1h/aula.

3. O plano de experimento: o cálculo da quantidade de experimentos a ser realizado depende do número de fatores escolhidos. O mesmo acontece com a escolha do plano de experimento. No caso desse trabalho, como os alunos sugeriram quatro alterações a realizar no modelo de avião Arumakan, optou-se pelo plano de experimento completo 2^4 , que equivale 16 experimentos. O nível desse plano é 2, ou seja, para cada fator existem duas possibilidades. Para que os alunos melhor entendessem, foi exposto um exemplo clássico, da combinação de roupas: “Uma pessoa possui duas camisas, duas bermudas e dois calçados, de quantas formas ela pode se vestir?”. Após essa elucidação, os alunos, através de cálculos, determinaram quantos experimentos haveriam de realizar, com base no princípio fundamental da contagem. Essa fase foi cumprida em 50 min.

4. Aula-treino: nesse momento, os alunos aprenderam a confeccionar o modelo de avião escolhido, pois se achou necessário que os alunos aprendessem a construir o modelo da melhor forma possível, minimizando as diferenças entre os aviões. A confecção dos aviões foi realizada em grupos de 3 a 4 alunos, com o auxílio do professor. Todos os grupos elegeram um líder, que fez mediação, em dados momentos, entre o professor e grupo. Antes do treino de confecção ser iniciado, foi falado sobre o desperdício de papel e fornecido aos grupos a quantidade exata de folhas. Na medida em que os alunos precisaram de mais folhas, registrou-se, em um lugar visível a todos, a quantidade excedente. No treino, foram utilizadas folhas de sulfite usadas. Como nessa etapa eles ainda estavam aprendendo a confeccionar o avião, houve um pequeno desperdício; por isso, essa questão não foi rigidamente considerada. Ainda nesse momento, realizou-se uma prévia do experimento. Os alunos foram para o ambiente extraclasse e efetuaram o lançamento dos aviões; usando réplicas do modelo inalteradas. A aula-treino serviu para os alunos: se adaptarem ao uso do lançador de aviões de papel (Apêndice H), utilizado para uniformizar os lançamentos; aprenderem a cronometrar o tempo de voo; praticarem o cálculo do tempo médio; e registrarem na tabela os dados obtidos (Apêndice B). Foram utilizados os cronômetros do *smartphone* dos alunos e sempre havia aparelhos o suficiente. O experimento ocorreu em local fechado, para não haver influência externa sobre o voo dos aviões, como, por exemplo, do vento. O local de lançamento foi em uma sala com dimensões iguais a 10 m de largura, 10 m de comprimento e 3 m de altura. Antes do lançamento, foi elaborado pelo professor um roteiro (Apêndice C), para que os grupos alternassem entre as funções de lançar, cronometrar e calcular a média. Todos participaram nas três funções. É importante lembrar que o experimento ficou a cargo dos alunos. Os líderes portavam o roteiro do experimento e eram os responsáveis por saber qual atividade seu grupo realizaria. O papel do professor foi o de mediar as relações entre os líderes e seus respectivos grupos. Quando necessário, o docente abordava conteúdos próprios da realização da atividade. Cada lançamento ocorria da seguinte forma: um aluno contava até três (em voz alta) e o avião era lançado; simultaneamente, disparavam-se os cronômetros. Quando o avião atingia o chão, esses cronômetros eram parados; por fim, calculava-se a média dos três tempos de voo. Um conteúdo necessário para realizar essa etapa foi a média aritmética. Os alunos foram orientados com antecedência sobre como calculá-la. Também foi dito sobre a forma correta de preenchimento da tabela com os dados obtidos. A tabela foi preparada previamente pelo professor. Essa foi uma etapa trabalhosa, demorou em média 3h/aulas.

5. Sorteio das alterações: no início dessa fase, assim como das outras, foi lembrado aos alunos o que se fez anteriormente, para dar consistência ao projeto. A partir daí, sorteou-se, no *software* Minitab, os 16 aviões com suas respectivas configurações (alterações). Os alunos presenciaram todo o processo, pois foi utilizado o projetor digital. Acompanharam, passo a passo, como foi realizado o sorteio no programa. Fornecidas as características do Planejamento de Experimentos ao programa, ele retornou uma tabela (Apêndice E) com as configurações, a numeração e a ordem de lançamento de cada avião. Em todas as aulas, os alunos foram lembrados sobre os fatores escolhidos (abreviados em siglas). Essa etapa durou, no máximo, 50 min.

6. Confeção dos aviões com alterações: essa foi uma etapa muito delicada e, por isso, realizada com muito cuidado e capricho. Foi dito aos alunos, da forma mais clara o possível, o que deveria ser feito e como deveria ser feito. Foi fornecido a eles um gabarito com o desenho e as dimensões do avião (Apêndice D), que mostrava onde deviam medir e cortar para reduzir o tamanho do modelo. Para as dobraduras no papel, foi incentivado o uso da régua, que permitia um friso melhor na dobradura. Para o uso da cola, foi orientado não usar muita, apenas um “fio de cola”. Os materiais necessários para essa aula foram: régua, tesoura, lápis, borracha, cola, papel sulfite e papel novo antílope. Buscou-se deixar claro que os aviões deveriam sair, o mais próximo o possível, com as mesmas medidas, o que justifica a realização da aula-treino. Em geral, demora-se 2h/aula para o cumprimento da confecção dos aviões com suas alterações.

7. Realização do experimento: como a aula-treino serviu de preparação para essa etapa, ela não foi complicada. Os passos foram os mesmos: 1) lançamento; 2) cronometragem; 3) cálculo da média e registro. Seguindo o roteiro estabelecido na aula-treino, os grupos alternaram nas tarefas: 1) um grupo lança o avião usando o dispositivo de lançamento, seguindo a ordem retornada pelo programa Minitab; 2) simultaneamente, outro grupo marca o tempo e para os cronômetros quando o avião atingir o chão; 3) o grupo responsável pelo cálculo da média dos tempos dos cronômetros o faz e registra em tabela. Após todos os lançamentos, foi feita uma réplica do experimento, para comparar os resultados de ambos os experimentos e verificar a coerência. Como os alunos já praticaram essa atividade anteriormente, o tempo poderia ter sido menor, mas foi o mesmo (3h/aulas).

8. Análise estatística dos dados: esse foi um dos momentos mais importante do projeto. Nessa fase, os alunos aprenderam grande parte dos conceitos estatísticos e probabilísticos. O tratamento dos dados foi feito em oito etapas:

- 1) Cálculo da média entre os resultados das duas réplicas do experimento;
- 2) Construção do gráfico de colunas da média obtida;
- 3) Organização dos dados em Rol estatístico;
- 4) Aplicação da Fórmula de Sturges para o cálculo da quantidade de intervalos de classe e seus respectivos valores, visando à construção da tabela de frequências;
- 5) Construção da tabela de frequências;
- 6) Construção do Histograma;
- 7) Cálculo dos ângulos e porcentagens para construção do gráfico de setores;
- 8) Construção do gráfico de setores.

Durante essas atividades também foram abordados conceitos como: números decimais, razão e proporção, arredondamento de decimais, regra de três simples, geometria plana, etc.

Para ensino dos conteúdos foi preparada uma apresentação em *PowerPoint*. Toda ela foi baseada em um único exemplo, um conjunto de dados de altura (cm) de uma turma hipotética. Para todas as oito etapas que os alunos cumpririam havia um exemplo com base nos dados da turma hipotética. Para a etapa do Rol, por exemplo, mostrou-se como ficaria o Rol dos dados de altura da turma. Para cada etapa havia um exemplo e uma tarefa. Verificado que todos os grupos houvessem completado a tarefa da etapa atual, lhes era exibida a correção na apresentação. Claro que isso foi previamente preparado.

Para a primeira etapa, os alunos fizeram o cálculo da média utilizando papel, lápis e borracha. Como eram 16 aviões e havia dois tempos de voo para cada avião (experimento e réplica), o grupo teria que calcular 16 médias. Eles dividiram essa tarefa entre eles, cada um fazia uma parte dos cálculos. Em todo o momento, monitorava-se a atividade dos alunos. O líder de cada grupo sempre era cobrado da participação dos outros integrantes na atividade.

A segunda etapa, do Rol, foi muito simples. Os alunos organizaram os dados em ordem crescente ou decrescente. Eles compararam números da ordem de centésimos (duas casas decimais). Na terceira etapa, construíram o gráfico de colunas com base nos dados do

Rol. Após o término do gráfico de colunas, deu-se início à preparação para construção da tabela de frequência (4ª etapa). Para isso, os alunos calcularam a quantidade de intervalos de classes e seus valores usando a equação de Sturges.

Na quinta etapa, com base nos dados obtidos anteriormente, os grupos construíram a tabela frequências. Terminada a tabela, deu-se início à construção do histograma (6ª etapa). Os alunos construíram o histograma olhando as informações da tabela de frequências. A penúltima tarefa foi obter os dados necessários para construção do gráfico de setores. Os alunos calcularam o valor dos ângulos e da porcentagem de cada frequência utilizando a regra da proporcionalidade (regra de três simples). Aprenderam a medir e construir ângulos utilizando o transferidor; e a traçar circunferências com o compasso, dado um raio (cm) qualquer. Por fim, aptos a construir o gráfico de setores, o fizeram com esmero. Essa fase durou 6h/aulas.

9) Conclusão e recomendação: terminadas as tarefas de análise estatística, toda a sala, como um único grupo, discutiu sobre qual alteração permitiu o avião voar mais tempo. Após muita discussão e argumentações, os alunos entraram em acordo. Para finalizar essa etapa, auxiliados pela professora de Língua Portuguesa, eles elaboraram um relatório científico sobre o experimento. Com base nos dados estatísticos e observações, relataram suas conclusões. Por fim, recomendaram a alteração que aperfeiçoa o tempo de voo do modelo escolhido por eles. Essa etapa durou 50 min.

Ao fim das atividades, aplicou-se novamente o TEPEP (como Pós-Teste), para verificar se a sequência didática teve influência estatística na mudança do nível de Proficiência em Estatística e Probabilidade dos alunos. Para auxiliar na realização de toda metodologia, descrita anteriormente, foram utilizados *softwares*, apresentados a seguir.

4.4 FERRAMENTAS TECNOLÓGICAS UTILIZADAS

Os *softwares* utilizados neste trabalho foram: *R*, Minitab e PowerPoint. O programa *R*¹ constitui um ambiente com linguagem para cálculos estatísticos e visualização de gráficos. É um programa GNU (100 % livre) com linguagem similar à linguagem *S*, desenvolvido na *Bell Laboratories* por John Chambers e outros. Há

---

¹ **The R Project for Statistical Computing.** Disponível em: <<https://www.r-project.org/>>. Acesso em: 23 jul. 2016.

diferenças, mas grande parte do código escrito para *S* é o mesmo em *R*. O programa possibilita: manipulação e armazenamento de dados; um conjunto de operadores para cálculos em arranjos, particularmente matrizes; uma grande, coerente e integrada coleção de ferramentas intermediárias para análise de dados; pacotes para análise e visualização de gráficos e linguagem simples e efetiva, com funções definidas pelos próprios usuários. Foi utilizado nesta pesquisa para a construção e análise de gráficos e realização de cálculos psicométricos através dos pacotes: *psych*, *TAM*, *CTT*, *polycor*, etc.

O *Minitab Statistical Software*² foi criado em 1972 por professores da Universidade Estadual da Pensilvânia, nos Estados Unidos. Seu objetivo era facilitar o ensino de estatística. Hoje, educadores em mais de 4000 faculdades e universidades usam o Minitab para ensinar análise de dados. É um programa pago. Neste projeto, utilizou-se a versão demonstrativa (17.2), válido para 30 dias, apenas para realizar o Planejamento de Experimentos. Para a criação de apresentação em *slides*, feita para exemplificar certos conceitos aos alunos, empregou-se o programa PowerPoint³ (2010), da empresa Microsoft.

---

² **Minitab**. 2016. Disponível em: <<https://www.minitab.com/pt-br/company/>>. Acesso em: 23 jul. 2016.

³ **PowerPoint**. 2016. Disponível em: <<https://products.office.com/pt-br/powerpoint>>. Acesso em: 23 jul. 2016.

5 RESULTADOS E DISCUSSÃO

Os resultados dessa pesquisa estão dispostos em duas partes. Na primeira, é abordada a validação do Teste de Proficiência em Estatística e Probabilidade (TEPEP), considerando os processos de validação adotados na metodologia, conforme descrito no item 4.2 (Figura 4). A segunda parte trata da análise da eficiência do projeto com os aviões de papel.

5.1 VALIDAÇÃO DO TESTE DE PROFICIÊNCIA EM ESTATÍSTICA E PROBABILIDADE

O primeiro critério de validade analisado no TEPEP foi o de validade de conteúdo (item 3.4). Tratou-se da submissão do teste a um júri composto por 5 especialistas. Os resultados do julgamento (Apêndice F) foram tratados por meio do Coeficiente de Validade de Conteúdo (CVC) (HERNANDEZ-NIETO, 2002). Os resultados são exibidos na Tabela 5.

Tabela 5 – Resultados de Coeficiente de Validade de Conteúdo (CVC) do julgamento do TEPEP.

Item	CL		PP		RT		MC	
	M_x	CVC_{ti}	M_x	CVC_{ti}	M_x	CVC_{ti}	M_x	CVC_{ti}
01	4,8	0,97	4,5	0,90	4,3	0,87	4,4	0,88
02	5,0	1,00	4,8	0,97	4,8	0,97	4,9	0,98
03	4,3	0,87	4,7	0,93	4,8	0,97	4,6	0,92
04	4,8	0,97	4,8	0,97	5,0	1,00	4,9	0,98
05	4,8	0,97	4,7	0,93	4,7	0,93	4,7	0,94
06	4,0	0,80	4,5	0,90	4,8	0,97	4,6	0,92
07	4,2	0,83	4,7	0,93	4,8	0,97	4,6	0,92
08	4,2	0,83	4,5	0,90	4,5	0,90	4,5	0,90
09	4,5	0,90	4,7	0,93	4,7	0,93	4,6	0,92
10	4,3	0,87	4,3	0,87	4,3	0,87	4,3	0,86
11	4,5	0,90	4,5	0,90	4,7	0,93	4,5	0,90
12	4,5	0,90	4,5	0,90	4,7	0,93	4,6	0,92
13	4,7	0,93	5,0	1,00	4,8	0,97	4,9	0,98
14	4,7	0,93	4,8	0,97	4,8	0,97	4,8	0,96
15	4,5	0,90	4,7	0,93	4,8	0,97	4,7	0,94
CVC_t	0,9044		0,9288		0,9423		0,92998	

Fonte: Próprio do autor;

Nota: CL=Clareza de Linguagem, PP= Pertinência Prática, RT=Relevância Teórica e MC= Média dos critérios.

Sobre a Clareza de Linguagem (CL), obteve-se um $CVC_t = 0,9044$, muito acima do valor mínimo desejado ($\geq 0,70$). Os itens 06, 07 e 08 apresentaram os CVC_{ti} com valores

mais baixos: 0,80, 0,83 e 0,83, respectivamente, ainda acima do ponto de corte. O item 08 foi mais criticado e foram feitas alterações em seu texto, sugeridas pelo júri, mas não houve necessidade de reavaliação do item, visto que os valores de CVC_{ti} foram significativos.

Tratando da Pertinência Prática (PP), o TEPEP expressou um $CVC_t = 0,9288$, muito satisfatório. O valor mais baixo de CVC_{ti} pertence ao item 10 (questão sobre probabilidade), que ainda assim permanece dentro do parâmetro estabelecido.

Sobre a Relevância Teórica (RT), os itens 01 e 10 foram os que apresentaram menor valor de CVC_{ti} , ambos com 0,87, ainda no limite. O valor do CVC_t encontrado foi de 0,9423, que está muito acima do ponto de corte.

Considerando a média entre as notas dos critérios de cada item (MC), obteve-se um CVC_t , para todo o teste, igual a 0,93, que é excelente, comparado ao limite.

O menor valor de CVC_{ti} foi 0,86, correspondente ao item 10. Em relação às médias (M_x) dos três critérios julgados e ao teste como um todo, nota-se que o teste foi avaliado acima da média, com menor média no item 06 (4,0), relacionada à Clareza de Linguagem.

Conforme os resultados de CVC, a validade de conteúdo foi atingida. Os 15 itens do TEPEP foram mantidos. Outro preceito analisado foi a validade de critério.

Segundo as orientações de Pasquali (2003), a validade de critério pode ser atingida correlacionando o desempenho dos alunos no teste com o desempenho dos mesmos sujeitos em atividade similar. Os dados utilizados na validade de critério foram os dos 17 alunos que participaram do projeto com os aviões de papel. Correlacionou-se o desempenho desses sujeitos no TEPEP, quando em Pós-Teste, com o desempenho acadêmico geral (Y) dos mesmos em seis disciplinas cursadas no primeiro bimestre de 2016. Para averiguar o grau de correlação entre os parâmetros comparados foi empregado o coeficiente de correlação de Pearson (RODGERS; NICEWANDER, 1988),

Na Tabela 6 são apresentados o desempenho dos 17 alunos no TEPEP (X) e o desempenho acadêmico geral (Y), ambos também em escala de porcentagem.

Tabela 6 – Desempenho dos alunos.

Aluno	X	Y	X(%)	Y(%)
01	8	6,50	0,65	0,53
02	10	7,25	0,73	0,67
03	13	5,42	0,54	0,87
04	11	8,50	0,85	0,73
05	15	9,75	0,98	1,00
06	6	7,00	0,70	0,40
07	11	7,33	0,73	0,73
08	10	7,00	0,70	0,67
09	7	6,67	0,67	0,47
10	12	7,00	0,70	0,80
11	4	6,17	0,62	0,27
12	8	6,92	0,69	0,53
13	7	4,33	0,43	0,47
14	14	8,92	0,89	0,93
15	7	6,33	0,63	0,47
16	10	5,50	0,55	0,67
17	5	5,08	0,51	0,33

Fonte: Próprio do autor;

Nota: X=Desempenho no TEPEP; Y=Desempenho Acadêmico.

Utilizando como critérios X (%) e Y (%), obteve-se um coeficiente de Pearson aproximado a 0,61, que especifica, segundo Hinkle, Wiersma e Jurs (1998, p. 120), uma correlação positiva moderada, ou seja, satisfatória. Determinado o valor de $r^2 = 0,3732$ e seguindo os critérios de Knoke e Mee (2002), para pesquisa social, pode-se afirmar que o valor é significativo e raramente é observado. Outro critério aplicado para a verificação da significância do coeficiente de correlação de Pearson foi o de Rosner (1995, p. 507), que testa a hipótese nula de que não há correlação entre a população que realizou o teste e a que participou do critério de comparação. Para isso, buscou-se satisfazer a condição:

$$r > \frac{t_{n-2, 0,975}}{\sqrt{n-2 + (t_{n-2, 0,975})^2}} \quad (\text{conforme eq. 7})$$

Assim, fazendo uso da tabela de distribuição t e conhecido o valor de r :

$$0,61 > \frac{t_{17-2, 0,975}}{\sqrt{17-2 + (t_{17-2, 0,975})^2}} = \frac{t_{15, 0,975}}{\sqrt{15 + (t_{15, 0,975})^2}} = \frac{2,1315}{\sqrt{15 + 2,1315^2}} \cong 0,48$$

Logo, a condição foi satisfeita e confirmada a significância r , rejeitando H_0 .

No gráfico de dispersão (Gráfico 5), observa-se, de forma mais clara, a correlação linear entre o Teste (X) e o critério (Y).

Gráfico 5 – Correlação entre o desempenho acadêmico dos alunos e o desempenho no TEPEP.

Fonte: Próprio do autor.

Nota-se que há linearidade entre os resultados do TEPEP e o desempenho acadêmico dos alunos. Dessa forma, o teste tem a capacidade de discriminar os alunos com maior desempenho dos alunos com menor desempenho, atingindo a validade de critério.

Conforme afirma Raymundo (2009), outro requisito básico para se construir um instrumento de avaliação confiável é a fidedignidade. Para isso, foi empregado o método da metade, que é estimado pela fórmula de Spearman-Brown (VIANNA, 1978). Considerando o resultado da aplicação do TEPEP aos 374 alunos de 9º ano tem-se:

$$r_{xx} = \frac{\frac{\sum XY}{N} - \bar{X} \bar{Y}}{S_X S_Y} = \frac{\frac{8333}{374} - 4,8983 \cdot 4,1149}{1,8351 \cdot 1,6921} \cong 0,68$$

Ou seja, para a metade do teste a fidedignidade é 0,68 e para todo o teste:

$$r_{(n)x} = \frac{nr_{xx}}{1 + (n-1)r_{xx}} \rightarrow r_{(2)x} = \frac{2 \cdot 0,68}{1 + (2-1) \cdot 0,68} \cong 0,81$$

A fidedignidade calculada foi 0,81. Isso significa que 81 % da variância nos escores do teste dependem da variância verdadeira e 19 % da variância dos erros.

O último critério de validade analisado foi o de construto, através da Análise Fatorial Exploratória (AFE), que demanda um pressuposto: a fatorabilidade da matriz de correlação dos dados. Para isso, aplicou-se o método de Kaiser-Meyer-Olkin (KMO); e o teste de esfericidade de Bartlett (KAISER, 1960, 1970; BARTLETT, 1950).

O programa estatístico R foi utilizado para a obtenção do valor de KMO, mas como os dados coletados na aplicação do TEPEP são categóricos, representativos (acerto = 1, erro = 0), fez-se uso de funções do programa para o computador assumir valores

numéricos 0 e 1 como lógicos, buscando determinar a matriz de correlação apropriada. Feito isso, obteve-se $KMO = 0,67$. Conforme Kaiser (1974), esse valor pode ser classificado como mediano e bem próximo de meritório, acima do mínimo determinado (0,50). Assim, existe a possibilidade da realização da AFE, para análise da validade de construto.

Para reafirmar a consistência interna dos dados, foi aplicado o teste de esfericidade de Bartlett (1950). Investigou-se a hipótese nula (H_0) de que não há correlação entre os itens do teste, ou seja, a matriz correlação é matriz identidade e o modelo fatorial é inadequado. Para rejeitar H_0 , o valor da estatística p (p -value) deve ser menor do que 0,05 (THOMPSON, 2004; TABACHNICK; FIDELL, 2013). Obteve-se no teste um valor de $\chi^2 = 2315,37$ e $p = 0$, rejeitando a hipótese nula, isso com significância igual a 95 % e 105 graus de liberdade, caracterizando correlação entre os itens.

Constatada a fatorabilidade da matriz de correlação, iniciou-se a análise da validade de construto, através da AFE. Conforme a sequência proposta por Willians, Brown e Osman (2010), extraíram-se primeiramente os fatores (componentes). Para auxiliar nessa tarefa, fez-se uso do *Screeplot* e da variância acumulada. No Gráfico 6, pode se ver o gráfico *Screeplot* em análise paralela, para oito componentes.

Gráfico 6 – *Screeplot* em análise paralela.

Fonte: Próprio do autor.

Obteve-se o *Screeplot* através da função *fa.parallel* do pacote *psych* (programa R), que calcula os autovalores tanto para a Análise das Componentes Principais (ACP) quanto para a Análise Fatorial (AF), fornecendo uma simulação para cada análise, ou seja, um parâmetro de comparação que auxilia na determinação dos componentes/fatores a

serem extraídos. Nota-se que os autovalores reais da ACP (CP atual), que antecedem a terceira componente, têm valor acima dos valores simulados (CP simu.); porém, a partir da terceira componente, os valores simulados superam os valores reais. Isso significa que a análise paralela sugere a extração apenas das componentes 1 e 2, que é coerente, pois, conforme sugere Cattell (1966), componentes/fatores com autovalores menores do que 1,0 devem ser descartados. Comparando os autovalores reais da AF com sua simulação, nota-se que a partir do 7º fator os valores reais são superados pelos simulados, indicando a retenção dos seis primeiros fatores. Porém, baseando-se na reta de corte ($y=1$), percebe-se que somente a primeira componente tem valor maior do que 1,0. Constatada essa divergência, recorreu-se à observação dos valores de variância acumulada. Para maior clareza da análise paralela do *Screeplot*, são exibidos, na Tabela 7, os valores dos autovalores reais e simulados, da matriz de correlação para oito componentes/fatores.

Tabela 7 – Análise das principais componentes e fatores.

Comp./Fator	CP atual	CP simu.	AF atual	AF simu.
1	5,409	1,359	4,756	0,547
2	1,531	1,284	0,844	0,301
3	1,163	1,216	0,472	0,241
4	1,044	1,166	0,287	0,191
5	0,915	1,110	0,215	0,133
6	0,837	1,068	0,135	0,091
7	0,709	1,023	0,047	0,047
8	0,674	0,987	-0,027	0,012

Fonte: Próprio do autor;

Nota: CP=Componentes Principais e AF=Análise Fatorial.

Para o emprego da porcentagem de variância acumulada, através da ACP, consideraram-se as 15 variáveis (componentes) e foram calculados seus autovalores a partir da matriz de correlação dos dados, utilizando a função *princomp* (pacote *pysch*), do programa R. Os resultados são exibidos na Tabela 8.

Tabela 8 – Variância acumulada segundo a Análise das Componentes Principais (ACP).

Componentes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Proporção de variância	0,36	0,10	0,08	0,07	0,06	0,06	0,05	0,04	0,04	0,04	0,03	0,03	0,02	0,01	0,01
Proporção acumulada	0,36	0,46	0,54	0,61	0,67	0,73	0,77	0,82	0,86	0,90	0,93	0,96	0,98	0,99	1,00

Fonte: Dados da pesquisa.

Nota-se que as seis primeiras componentes possuem variância acumulada de 73 %. Segundo O'Rourke (2012), quando agrupados, os autovalores devem ter variância acumulada mínima de 70 % a 80 % e sozinhos, de 5 % a 10 %. Individualmente, acima da 7ª componente, os autovalores apresentam uma proporção de variância de 4 % a 1 %, não respeitando o critério. Hair et al. (2009) afirmam que a variância acumulada deve ter no mínimo 60 %. Os autovalores dentro deste limite estão acima da 3ª componente.

Com base na interpretação do *Screeplot* e dos resultados da variância acumulada, ficou estabelecida a extração das 4 primeiras componentes. Na Tabela 9, estão os valores das cargas das componentes para as 15 variáveis e suas comunalidades. A Figura 5 apresenta, com mais clareza, a relação das componentes, suas cargas e os itens do teste.

Tabela 9 – Cargas das componentes.

Item	CP.1	CP.2	CP.3	CP.4	Com.
01	0,53	-0,37	-0,46	-0,30	0,72
02	0,50	0,44	0,18	-0,06	0,48
03	0,57	0,42	0,23	-0,35	0,67
04	0,47	0,25	0,32	0,27	0,46
05	0,59	-0,31	0,31	0,21	0,58
06	0,49	-0,50	0,38	-0,30	0,72
07	0,70	-0,10	0,25	-0,17	0,60
08	0,65	0,49	-0,02	-0,09	0,67
09	0,63	0,38	0,37	-0,18	0,71
10	0,58	-0,30	0,33	0,26	0,61
11	0,61	0,09	0,03	0,23	0,43
12	0,62	0,19	-0,29	0,31	0,60
13	0,64	0,26	-0,27	-0,39	0,70
14	0,60	-0,12	0,20	0,34	0,53
15	0,64	0,08	-0,38	0,35	0,68

Fonte: Próprio do autor;

Nota: CP.=Componente e Com.=Comunalidade.

Figura 5 – Distribuição das cargas das componentes sobre os itens.

Fonte: Próprio do autor;

Nota: PC=Principal Components.

Observando os resultados das cargas das componentes principais, nota-se que a primeira componente (CP.1) possui cargas mais altas, todas acima de 0,40. As componentes de 2 a 4 não apresentam cargas tão altas. Seguindo a sugestão de Guadagnoli e Velice (1988), que consideram o tamanho da amostra, foi retida para interpretação somente CP.1, pois cinco de suas cargas têm valor maior que 0,60. Já CP.2, apresenta menos de 10 cargas com valor acima de 0,40; e o restante das componentes possuem cargas baixas. Apesar de o item 6 apresentar maior carga sobre CP.2, o valor da diferença

entre sua carga em CP.1 (0,49) e CP.2 (0,50) é insignificante, podendo-se considerar, no contexto geral dos resultados, que o item 6 se relaciona à CP.1; e não há necessidade de descarte do item, pois CP.2 não possui cargas altas para ser alvo de interpretação. Em relação à comunalidade (Com.), todos os itens apresentam valores satisfatórios, com maioria entre 0,40 e 0,70, conforme Osborne e Costello (2009).

Os resultados da ACP indicaram que o TEPEP é unidimensional, avalia somente o construto Proficiência em Estatística e Probabilidade, que corresponde à primeira componente observada (CP.1). Constatada essa característica, cumpriu-se a premissa para aplicação da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI).

5.2 TEORIA CLÁSSICA DOS TESTES

O primeiro conceito verificado no TEPEP, a partir da Teoria Clássica dos Testes (TCT), foi o de assimetria, ou seja, o desvio ou o afastamento da simetria (TOLEDO; OVALLE; 1983). De acordo com Urbina (2007), quando a distribuição dos escores é negativamente assimétrica, pode-se dizer que o teste é fácil; dada a mesma população, se a distribuição for assimetricamente positiva, o teste é difícil. No Gráfico 7, é apresentado o histograma da distribuição dos escores dos 374 alunos que realizaram o TEPEP.

Gráfico 7 – Distribuição normal dos escores dos 374 alunos.

Fonte: Próprio do autor.

Através dos valores obtidos da média (9,01), moda (9,0) e mediana (9,0), a distribuição dos escores dos alunos foi classificada como simétrica, conforme sugerido por

Toledo e Ovalle (1983). Outro critério utilizado para confirmar a simetria dos dados foi o coeficiente de assimetria de Karl Pearson (α_3) (TOLEDO; OVALLE; 1983). Nesse caso, a simetria é comprovada quando o valor do coeficiente for igual a zero. O resultado obtido, calculado no *software* R, foi $\alpha_3 = 0,05 \cong 0$, classificando a distribuição como simétrica.

Aplicou-se a “Regra 27”, de Kelley, para análise do índice (D) de discriminação dos itens. Segundo seus escores, os alunos são divididos proporcionalmente em: Grupo Superior (GS) – alunos com 27 % das melhores notas; Grupo Inferior (GI) – alunos com 27 % das piores notas; e os 46 % restantes – alunos com notas próximas da média, aqui chamado de Grupo Mediano (GM). O cálculo das notas que diferem os melhores alunos (n_2) dos piores (n_1) foi realizado com base na distribuição normal padrão (Figura 6).

Figura 6 – Divisão dos grupos segundo a regra de Kelley, paralelo à distribuição normal padrão.

Fonte: Próprio do autor;

Nota: Média=9,01 e desvio padrão=3,24.

Através do cálculo da estatística z tem-se:

$$P(n \geq n_2) = 0,27 \Rightarrow P\left(z \geq \frac{n_2 - \mu}{\sigma}\right) = 0,27 \Rightarrow P\left(z_2 \geq \frac{n_2 - 9,01}{3,24}\right) = 0,27 \Rightarrow$$

$$\Rightarrow 0,50 - P\left(z_2 \geq \frac{n_2 - 9,01}{3,24}\right) = 0,23 \Rightarrow \frac{n_2 - 9,01}{3,24} = 0,61 \Rightarrow n_2 = 0,61 \cdot 3,24 + 9,01 \cong 11$$

$$\text{Para } n_1: P(n \leq n_1) = 0,27 \Rightarrow P\left(z \leq \frac{n_1 - \mu}{\sigma}\right) = 0,27 \Rightarrow P\left(z_1 \geq -\left(\frac{n_1 - 9,01}{3,24}\right)\right) = 0,27$$

$$\Rightarrow 0,50 - P\left(z_1 \geq -\left(\frac{n_1 - 9,01}{3,24}\right)\right) = 0,23 \Rightarrow -\left(\frac{n_1 - 9,01}{3,24}\right) = 0,61 \Rightarrow$$

$$\Rightarrow n_1 = (-0,61) \cdot 3,24 + 9,01 = 7,03 \cong 7$$

Determinados n_1 e n_2 , caracterizaram-se GS, GI e GM, conforme Kelley (1939):

Gráfico 8 – Divisão dos grupos inferior e superior.

Fonte: Próprio do autor.

Dada a classificação dos grupos, foram organizados seus escores e então calculados os índices de dificuldade e discriminação de cada item do TEPEP; a dificuldade de um item é representada pela soma da porcentagem total de acertos; maior a porcentagem, mais fácil é o item. No Quadro 3 são apresentados os resultados para os três grupos.

Quadro 3 – Desempenho dos grupos superior, médio e inferior, segundo a TCT.

GS (27%)	Item	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	acertos	118	115	118	107	75	88	101	115	113	83	92	116	118	91	82
	Erros	2	5	2	13	45	32	19	5	7	37	28	4	2	29	38
	Total	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
acertos(%)		98,3	95,8	98,3	89,2	62,5	73,3	84,2	95,8	94,2	69,2	76,7	96,7	98,3	75,8	68,3
GM (46%)	Item	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	acertos	121	109	119	82	33	50	66	114	85	24	52	93	111	29	40
	Erros	4	16	6	43	92	75	59	11	40	101	73	32	14	96	85
	Total	125	125	125	125	125	125	125	125	125	125	125	125	125	125	125
acertos(%)		96,8	87,2	95,2	65,6	26,4	40,0	52,8	91,2	68,0	19,2	41,6	74,4	88,8	23,2	32,0
GI (27%)	Item	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	acertos	108	76	90	45	19	36	18	62	34	20	23	63	75	21	21
	Erros	21	53	39	84	110	93	111	67	95	109	106	66	54	108	108
	Total	129	129	129	129	129	129	129	129	129	129	129	129	129	129	129
acertos(%)		83,7	58,9	69,8	34,9	14,7	27,9	14,0	48,1	26,4	15,5	17,8	48,8	58,1	16,3	16,3
Total de acertos		347	300	327	234	127	174	185	291	232	127	167	272	304	141	143
Total de erros		27	74	47	140	247	200	189	83	142	247	207	102	70	233	231
Total de acertos (%)		92,8	80,2	87,4	62,6	34,0	46,5	49,5	77,8	62,0	34,0	44,7	72,7	81,3	37,7	38,2
TOTAL		374	374	374	374	374	374	374	374	374	374	374	374	374	374	374
Discriminação		14,6	36,9	28,6	54,3	47,8	45,4	70,2	47,8	67,8	53,7	58,8	47,8	40,2	59,6	52,1

Fonte: Próprio do autor.

Observando o Total de acertos (%), pode-se notar que os itens 01, 02, 03, 08, 12 e 13 foram os mais fáceis, todos acima de 70 %. Sua maioria demanda interpretação de texto, de gráficos e tabelas. Desses itens, os que apresentaram discriminação (D) baixa foram: 01 (14,6) e 03 (28,6). Conforme critérios para o índice D, definidos por Ebel e Frisbie (1991), o item 01 é fraco e precisa de revisão; o item 03 está à margem do desejado e precisa de alguma revisão. Por outro lado, exceto o item 02, com discriminação razoável (36,9), todos os itens apresentaram discriminação muito boa ($> 0,40$). Os itens mais discriminativos foram: 07, 09, 11 e 14, que demandaram raciocínio elaborado e cálculo. De todo o teste, os itens considerados difíceis foram: 05 e 10, sobre média aritmética e probabilidade, respectivamente. Outro método aplicado para calcular a discriminação dos itens foi o Teste T (PASQUALI, 2003). Na Tabela 10 é apresentado esse resultado para os 15 itens do TEPEP, em ordem decrescente.

Tabela 10 – Índices de discriminação dos itens segundo os resultados do Teste T.

Item	07	09	14	11	04	10	15	12	08	05	13	06	02	03	01
D	43,2	37,8	28,7	28,0	24,6	23,5	22,3	21,4	21,3	19,7	18,6	17,9	16,9	15,2	12,0

Fonte: Próprio do autor.

É relevante notar que, os resultados do Teste T, comparados aos índices de discriminação do método da proporção, em relação à ordem, são similares. Os itens 07, 09, 14 e 11 continuam mais discriminativos e os itens 01, 03, 02 menos discriminativos.

Também foi empregada a correlação ponto-bisserial (PASQUALI, 2003, p. 135), para obtenção do índice D. Os resultados são apresentados na Tabela 11.

Tabela 11 – Índices de discriminação dos itens segundo a correlação ponto-bisserial.

Item	07	09	08	14	11	05	10	15	12	13	04	03	02	06	01
D	0,48	0,47	0,43	0,40	0,40	0,39	0,39	0,36	0,36	0,35	0,34	0,34	0,33	0,29	0,17

Fonte: Próprio do autor.

Apesar de alguns itens, antes menos discriminativos passarem a ter maior valor, a ordem ainda é similar: itens 07, 09 mais discriminativos e 01, 06, 02 e 03 menos discriminativos; revelando que os métodos aplicados são eficazes. Na próxima sessão serão discutidos os resultados para a Teoria de Resposta ao Item.

5.3 TEORIA DE RESPOSTA AO ITEM

Adotou-se, na Teoria de Resposta ao Item (TRI), o modelo logístico de dois parâmetros. Para verificar sua adequação aos dados coletados ao modelo, além de se verificar a unidimensionalidade, aplicou-se o Teste de Qui-Quadrado (χ^2 , equação 17, item 3.7) e foi construído o gráfico *Scatterplot*.

Para o Teste de Qui-Quadrado, considerou-se um nível de significância $\alpha = 0,05$ e 373 graus de liberdade, que corresponde a um valor crítico de Qui-Quadrado igual a 419,03 (obtido através da tabela de Qui-Quadrado). Como o valor calculado ($\chi^2 = 264,92$) é menor do que o valor crítico, aceitou-se, a hipótese nula, de que há consistência dos dados coletados com o modelo (HAMBLETON; SWAMINATHAN; ROGERS, 1991).

O *Scatterplot* (Gráfico 9) foi construído com base nos escores dos 374 sujeitos contra seus escores estimados pelo modelo.

Gráfico 9 – *Scatterplot* dos dados coletados contra os dados estimados pelo modelo.

Fonte: Próprio do autor.

Observa-se que existe linearidade entre os dados estimados pelo modelo e os dados coletados. Quanto maior a linearidade, mais real é a estimativa (PASQUALI, 2003; LORD, 1974). No Gráfico 10, é apresentada a Curva Característica do Teste (CCT) (BAKER, 2001; MACDONALD; PAUNONEN, 2002; PASQUALI, 2003), construída com base no resultado dos 374 alunos submetidos ao TEPEP.

Gráfico 10 – Curva Característica do TEPEP.

Fonte: Próprio do autor.

Nota-se que alunos com proficiência (θ) abaixo de -2 terão dificuldade para realizar o teste. Por outro lado, sujeitos com habilidade mediana ($\theta = 0$) acertariam pelo menos 9 questões e alunos com $\theta > 2$ terão um escore alto. No histograma (Gráfico 11), é exibida a distribuição da frequência dos 374 alunos, conforme seus níveis de proficiência.

Gráfico 11 – Histograma da proficiência em estatística e probabilidade (θ), dos 374 alunos.

Fonte: Próprio do autor.

De acordo com os dados do histograma, poucos alunos apresentaram dificuldade para realizar o TEPEP. A maioria demonstrou habilidade acima de -2,2. A concentração dos alunos com proficiência mediana encontra-se entre -1,3 e 1,3 e acima de 1,3, o grupo com maior habilidade. Houve mais alunos no nível avançado do que com baixo rendimento. Para obter desempenho mínimo no teste não se requisita muita proficiência,

mas para atingir alta pontuação é preciso maior habilidade, ou seja, o teste é discriminativo.

Os índices de dificuldade (*b*) e discriminação (*a*) (item 3.7) dos itens são expressos na Tabela 12, assim como seus valores de *infit* e *outfit*, todos em ordem decrescente.

Tabela 12 – Índices do modelo de dois parâmetros.

Item	b	Item	A	Item	Infit	Item	Outfit
5	0,812	8	1,953	15	1,0061	14	1,262
10	0,811	9	1,786	4	1,0052	3	1,084
14	0,615	3	1,781	8	1,0034	13	1,072
15	0,566	7	1,574	7	1,0029	1	1,063
11	0,247	13	1,461	10	1,0028	5	1,057
6	0,148	2	1,241	2	1,0004	11	1,013
7	-0,019	12	1,210	3	0,9993	12	1,005
4	-0,635	14	1,182	6	0,9987	6	1,002
9	-0,804	11	1,175	12	0,9941	2	1,000
12	-1,262	5	1,128	11	0,9938	10	0,986
2	-1,784	10	1,123	9	0,9935	9	0,980
13	-1,993	15	1,031	13	0,9866	7	0,978
8	-2,002	4	1,007	5	0,9835	15	0,956
1	-2,799	1	0,795	1	0,9805	8	0,936
3	-2,818	6	0,780	14	0,9761	4	0,919

Fonte: Próprio do autor;

Nota: b=dificuldade; a=discriminação.

De acordo com os dados da tabela, pode-se ver que os itens mais difíceis foram o 05, 10, 14 e 15 e os mais fáceis: 03, 01, 08 e 13. Comparando com os resultados da TCT, nota-se que, considerando o índice de dificuldade, os resultados estão de acordo.

Em relação à discriminação, os itens com maior valor foram os itens 08, 09, 03 e 07; e os menos discriminativos: 06, 01, 04 e 15. Para os resultados da TCT (Tabela 10), os menos discriminativos foram o 01, 03 e 02 e os mais discriminativos: 07, 09, 14 e 11. Para o índice de discriminação, as teorias se distinguem. O item 03, por exemplo, na TCT é menos discriminativo e na TRI, mais. Como as diferenças entre os modelos matemáticos das teorias são diferentes e se baseiam em conceitos diferentes, não se espera que sempre os resultados se comprovem. Em todo o caso, a TRI apresenta mais recursos e deve ser aplicada para refinar a análise dos itens (SARTES; SOUZA-FORMIGONI, 2013).

Sobre *infit* e *outfit*, seus valores estão próximos ao esperado (1,0) (LINACRE, 2012, p. 320; HUTZ; BANDEIRA; TRENTINI, 2015); ou seja, o modelo se ajusta bem aos itens do teste e não há muita discrepância nos dados (respostas inesperadas).

Finalizadas as análises da TCT e TRI, serão apresentados a seguir os resultados do desempenho dos alunos no TEPEP após a participação no projeto.

5.4 RESULTADOS DO PROJETO

O primeiro resultado obtido no projeto foi o da avaliação diagnóstica (Pré-Teste), feita através do Teste de Proficiência em Estatística e Probabilidade (TEPEP). No Gráfico 12 é apresentado o grau de conhecimento prévio dos 17 alunos participantes do projeto.

Gráfico 12 – Resultado do Pré-Teste.

Fonte: Próprio do autor.

Pode-se notar que 9 alunos ficaram abaixo da média (8,23) e a menor nota foi do aluno 13 (A13). O ponto preocupante foi que mais da metade da turma apresentou notas inferiores a 8,23. Isso significa que possuíam pouco conhecimento estatístico e probabilístico antes de participarem do projeto. Dos alunos que tiveram pontuação acima da média, apenas 2 (A03 e A05) demonstraram maior nível de conhecimento. O restante (6 alunos) conseguiu um escore próximo de 10. Com o objetivo de esclarecer o desempenho

dos alunos com escore inferior à média, suas pontuações nas questões são exibidas na Tabela 13.

Tabela 13 – Escores nos 15 itens dos alunos com desempenho abaixo da média.

	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12	I13	I14	I15
A01	1	0	0	0	0	1	1	1	0	0	0	1	1	0	0
A04	0	0	1	1	0	0	1	1	0	0	0	1	1	0	0
A06	1	1	1	1	0	0	1	0	0	1	1	0	0	0	0
A11	1	0	1	0	0	0	0	1	0	0	0	0	1	0	1
A12	1	1	1	1	0	1	0	1	0	0	0	1	0	0	0
A13	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0
A14	1	0	1	0	0	1	1	1	0	0	1	1	1	0	0
A16	1	1	0	0	0	1	0	1	0	0	0	1	1	0	0
A17	1	0	1	1	0	0	0	0	0	0	1	1	1	0	0
Esc.	7	3	6	4	0	5	5	6	0	2	3	6	6	0	1

Fonte: Próprio do autor;

Nota: 0=erro e 1=acerto.

Percebe-se que os alunos com notas abaixo da média tiveram pior escore (Esc.) nos itens 02, 04, 05, 09, 10, 11, 14 e 15. Desses itens, os que não computaram acertos, ou quase nenhum acerto, foram o 05, 09, 10, 14 e 15 (itens críticos). A média do grupo de alunos apresentado na tabela foi de 6,0. De acordo com a Curva Característica do Teste (CCT, p. 39), a proficiência (θ) em Estatística e Probabilidade do grupo está próxima à -1,5. Com base nessas informações e buscando entender o comportamento dos alunos em suas respostas, analisou-se a Curva Característica (CCI) de cada item crítico, assim como o padrão de resposta do grupo às alternativas desses itens (descritores – Tabela 14).

Tabela 14 – Descritores dos itens.

Itens	Alternativas			
	A	B	C	D
I05	0	0	9	0
I09	0	3	2	4
I10	5	0	2	2
I14	3	4	2	0
I15	1	1	5	2

Fonte: Próprio do autor.

O primeiro item menos pontuado foi o 05. Sua CCI é exposta no Gráfico 13.

Gráfico 13 – Curva Característica do Item 05.

Fonte: Próprio do autor.

Como a proficiência do grupo com baixo desempenho foi de -1,5, a probabilidade esperada dos alunos acertarem o item é menor do que 10 %. Segundo o índice de dificuldade (b) obtido na Teoria de Resposta ao Item (TRI), esse é o item mais difícil do TEPEP, o que se comprova na Teoria Clássica dos Testes (TCT). Observando a Tabela 14, pode-se ver que todos os alunos responderam o item na alternativa C. A questão é sobre média aritmética (vide Apêndice A) e os alunos, ao olharem para a tabela da questão e notarem que o número cinco mais se repete, acreditaram que a média teria o mesmo valor, ou seja, não conheciam o conceito e, talvez inconscientemente ou intuitivamente, confundiram com outra medida de tendência central, a moda.

O segundo item crítico foi o 09:

Gráfico 14 – Curva Característica do Item 09.

Fonte: Próprio do autor.

Conforme o índice de discriminação (α) verificado na TRI, o item 09 é o segundo mais discriminativo do TEPEP; o mesmo foi constatado na TCT. Nessa questão, considerada a proficiência do grupo (-1,5), os alunos teriam probabilidade de acerto esperada entre 10 % e 20 %. Essa questão trata da frequência relativa, um conceito, geralmente, pouco abordado nas aulas destinadas ao ensino de Estatística e Probabilidade. A maior parte dos alunos, usando o raciocínio dedutivo, acreditou que a alternativa correta fosse a D, pois o enunciado da questão os leva a pensar dessa forma, ou seja, “o dado é lançado 20 vezes e o número procurado é o 5”, então a resposta induzida seria 5/20. O aluno, por falta de prática e significado do conceito de frequência, não soube responder ao item corretamente.

O próximo item menos pontuado foi o 10:

Gráfico 15 – Curva Característica do Item 10.

Fonte: Próprio do autor.

De acordo com a TRI e TCT, o item 10 é o segundo mais difícil do TEPEP. É uma questão que trata sobre a probabilidade da ocorrência de um evento. Observando o Gráfico 15 nota-se que, para uma proficiência próxima a -1,5, a probabilidade de acerto esperada seria menor do que 10 %. Já um aluno de desempenho mediano, com proficiência próxima a zero, teria 30% de chance de responder corretamente à questão. É interessante ressaltar que a maior parte dos alunos respondeu na alternativa A. Um suposto raciocínio para explicar essa resposta é: “como são 3 tonalidades cintilantes, relacionaram 3 com 30 %”. Por falta de conhecimento probabilístico, talvez, os alunos acreditaram que o melhor “caminho” seria esse.

O penúltimo item crítico foi o 14:

Gráfico 16 – Curva Característica do Item 14.

Fonte: Próprio do autor.

Esse é o terceiro item mais difícil, segundo a TRI e TCT. A probabilidade esperada de acerto do grupo de menor desempenho está abaixo de 10 %. Assim como o item 05, o item 14 demanda o conhecimento de média aritmética. A resposta dos alunos se concentrou na alternativa B. É interessante ressaltar que, ao procurar uma lógica para essa resposta, foi indagado aos alunos o modo como eles pensaram para resolver a questão, e o retorno foi o seguinte: “Nós chutamos o 15, pois foi o único número que não apareceu entre os números da pergunta”. Deve-se lembrar que, o modelo da TRI adotado não calcula o índice de “chute” do item (*guess*), para isso, seria necessário utilizar o modelo logístico de 3 parâmetros da TRI, que não se ajustou bem aos dados da pesquisa.

O último item menos pontuado foi o item 15:

Gráfico 17 – Curva Característica do Item 15.

Fonte: Próprio do autor.

Segundo a TCT e TRI, o item 15 é o quarto mais difícil. Apenas dois alunos responderam corretamente à questão. Nesse item, a chance de acerto do grupo de menor desempenho está próxima de 10 %. A questão trata sobre moda e mediana. Em geral, são conceitos estatísticos pouco abordados e não se esperava um resultado ótimo. A maioria dos alunos do grupo optou pela alternativa C, sendo que a correta seria a B. Não foi percebida uma lógica para essa resposta; acredita-se que seja devido ao acaso. Nota-se na CCI do item que, para ocorrer um acerto, o grau de proficiência em Estatística e Probabilidade do sujeito deve ser elevado; ou o aluno sabe, ou teria de recorrer ao “chute”. As CCI de todos os outros itens são disponibilizadas no Apêndice G.

Em suma, o grupo com menor desempenho foi diagnosticado com dificuldade nos seguintes conceitos: medidas de tendência central, frequência, probabilidade e porcentagem. De acordo com Goodsell et al. (1992, p. 11), o conhecimento prévio dos alunos deve ser considerado para se criar um novo saber, pois a aprendizagem ativa é um processo construtivo. Após o Pré-Teste, iniciaram-se às atividades do projeto.

A aplicação do projeto consistiu em 9 etapas, conforme descrito no item 4.3 deste trabalho. A seguir, serão apresentados os resultados das atividades realizadas com os alunos, registrados durante o cumprimento das etapas.

Na primeira etapa da sequência didática, de introdução ao projeto, buscou-se criar um ambiente motivador. Ao exibir, na primeira vez, os aviões de papel e perguntar “*Quais alterações podem ser feitas em um modelo de avião de papel para que ele permaneça mais tempo voando?*”, observou-se entusiasmo nos alunos. Os conceitos da Aprendizagem colaborativa afirmam que, ao invés de se iniciar uma aula com fatos e ideias, direcionando para a aplicação, as atividades devem começar com problemas (GOODSELL et al., 1992, p. 12).

Outra questão importante foi a do voto para escolha do modelo de avião. Todos respeitaram a opinião do colega, caracterizando senso de responsabilidade cívica (idem, p. 14). Bonwell e Eison (1991) afirmam que, na aprendizagem ativa, dá-se menos ênfase à transmissão de informação e mais atenção ao desenvolvimento das habilidades dos alunos.

Durante a realização do *Brainstorm*, para listar as alterações a serem feitas no modelo escolhido, os alunos demonstraram respeito às ideias de seus colegas, fossem elas sem fundamento ou não. Somente ao fim de todo o processo, eles discutiram sobre quais

alterações realizariam. Essa atividade apontou para o senso de cooperação e trabalho em equipe, pois houve aceitação de todas as vozes do grupo, além de ser um processo inerentemente social (GOODSELL et al., 1992, p. 12-13). Após a discussão e argumentação, a sala como um todo concluiu que realizaria quatro alterações no modelo de avião. A aprendizagem ativa proporciona essas situações, onde os estudantes estão engajados em atividades como ler, escrever e *discutir* (BONWELL; EISON, 1991).

No decorrer da aula-treino, para a confecção e realização do experimento, os alunos se envolveram e aprenderam a trabalhar em equipe. Ampliaram o senso de organização coletiva. Na fase de confecção, uns auxiliaram os outros. No experimento prévio, os líderes, escolhidos pelos próprios elementos do grupo, aprimoraram a capacidade de liderança. Cultivar o trabalho em equipe e a habilidade de liderança é meta legítima e valiosa da sala de aula e não extracurricular (GOODSELL et al., 1992, p. 13). As etapas 6 e 7 apresentaram resultados similares, característicos da aprendizagem ativa e colaborativa.

As etapas onde os alunos tiveram uma participação menos ativa foram a 3 e 5. Na terceira etapa, de cálculo do total de rodadas do experimento, tiveram participação parcial: após exposição do “exemplo da combinação de roupas”, os alunos calcularam o total de experimentos a se fazer, utilizando o princípio fundamental da contagem.

Na etapa 5, do sorteio das alterações dos aviões, feito no *software* Minitab, os alunos comunicavam as informações do experimento ao professor e ele as fornecia ao programa, para que o *software* retornasse todo o plano do experimento. Vale lembrar que aulas colaborativas não vêm extinguir as aulas expositivas, mas os professores devem se portar mais como projetistas das experiências intelectuais de seus alunos (GOODSELL et al., 1992, p. 11).

A etapa 8, de análise dos dados obtidos no experimento, apresenta resultados dos conteúdos estatísticos aplicados pelos alunos. A primeira parte dessa etapa está relacionada ao cálculo da média entre os resultados do experimento e da réplica do experimento.

Na Figura 7, são exibidos os cálculos da média do tempo de voo dos aviões 02, 03 e 04.

Figura 7 – Cálculo de média aritmética feito pelos alunos.

Fonte: Próprio do autor.

Para o avião 02, o tempo no primeiro experimento foi igual a 1,30 s e na réplica 1,81 s. O aluno somou os tempos e dividiu pelo número de tempos. Fez o mesmo para os outros aviões. Então construíram um gráfico de colunas com os resultados obtidos (Gráfico 18).

Gráfico 18 – Gráfico de colunas construído pelos alunos.

Fonte: Próprio do autor.

Nota-se que a escala do eixo da média do tempo de voo está além do necessário. É um erro comum, que foi corrigido posteriormente. Finalizado o gráfico de colunas, os alunos organizaram os dados brutos em Rol estatístico, preparando-se para construção da tabela de frequências. Na Figura 8, é exibido o Rol construído pelos alunos.

Figura 8 – Rol organizado pelos alunos.

Dados Brutos									
1,14	1,55	1,52	1,45	1,46	1,15	1,31	1,44	1,33	
1,36	1,26	0,82	1,00	1,48	1,84	1,63			

Rol									
0,82	1,00	1,14	1,26	1,31	1,33	1,36	1,44	1,48	
1,52	1,55	1,63	1,75	1,76	1,84				

Fonte: Próprio do autor.

Os dados foram organizados corretamente. Haviam se esquecido do 1,15s no Rol, mas perceberam. Antes de construírem a tabela de frequências, os alunos tiveram de realizar o cálculo da quantidade de intervalos de classe e seus respectivos valores. Para isso, utilizaram a fórmula de Sturges (Figura 9.a). Aplicaram a fórmula e verificaram cinco intervalos de classe iguais a 0,2. Calcularam o logaritmo de n utilizando a calculadora de seus *smartphones*. Após isso, construíram a tabela de frequências (Figura 9.b).

Figura 9 – Aplicação da fórmula de Sturges e construção da tabela de frequências pelos alunos.

a) Fórmula de Sturges		b) Tabela de Frequência	
$K = 1 + 3,3 \cdot \log(n)$		Tempo de voo	Frequência
$n = 16$		0,81-1,00	1
$K = 1 + 3,3 \cdot \log(16)$		1,00-1,2	3
$K = 1 + 3,3 \cdot 1,20$		1,21-1,4	4
$K = 1 + 3,96$		1,41-1,6	4
$K = 4,96 = 5$		1,61-1,8	3
$IC = \frac{1,84 - 0,82}{5} = 1,02 = 0,20 \dots = 0,2$		1,81-2,00	1
5 5		Total	16

Fonte: Próprio do autor;

Nota: a) Fórmula de Sturges, b) Tabela de Frequências.

A tabela de frequências foi elaborada corretamente. Com base nos resultados dessa tarefa, os alunos construíram o histograma (Gráfico 19).

Gráfico 19 – Histograma construído pelos alunos.

Fonte: Próprio do autor.

Conseguiram construir o histograma com o título um pouco fora do lugar ideal, mas as informações transmitidas estão todas corretas. Acabada essa etapa, os alunos começaram a preparar a construção do gráfico de setores. Para isso, calcularam a porcentagem e o ângulo correspondentes a cada frequência. O resultado dessas operações, para a frequência 3, é exibido na Figura 10.

Figura 10 – Cálculo de regra de três simples para porcentagem e ângulo, feito pelos alunos.

$$\frac{16}{3} = \frac{100\%}{x}$$

$$16 \cdot x = 100 \cdot 3$$

$$16 \cdot x = 300$$

$$x = \frac{300}{16}$$

$$x = 18,75\%$$

$$\frac{16}{3} = \frac{360^\circ}{x}$$

$$16 \cdot x = 360 \cdot 3$$

$$16 \cdot x = 1080$$

$$x = \frac{1080}{16}$$

$$x = 67,5^\circ$$

Fonte: Próprio do autor.

Aplicando a regra de três simples, os alunos fizeram corretamente os cálculos. Porém, antes de começarem a construir o gráfico de setores, foi necessário que aprendessem a utilizar o compasso e transferidor. Na figura 11, são exibidas duas imagens dos alunos realizando essas tarefas.

Figura 11 – Alunos utilizando compasso e transferidor.

Fonte: Próprio do autor.

Durante essa atividade, notou-se que muitos alunos, mesmo já no 8º ano, nunca haviam utilizado compasso ou transferidor. Após aprenderem a manusear os instrumentos, construíram o gráfico de setores (Gráfico 20).

Gráfico 20– Gráfico de setores construído pelos alunos.

Fonte: Próprio do autor.

Na etapa 9, de recomendação de um curso de ação, os alunos elaboraram um relatório científico para concluir a atividade, ou seja, determinar quais alterações fizeram o avião voar mais tempo.

Na Figura 12, é exibida a conclusão de um dos grupos da classe.

Figura 12 – Conclusão de um dos grupos do projeto sobre o experimento.

Fonte: Próprio do autor.

Os alunos, ao concluírem, classificaram os aviões em dois grupos, os que voaram mais e os que voaram menos tempo. Discriminaram as alterações de cada um dos grupos, indicando as otimizações para o modelo escolhido pela turma. Terminadas todas as etapas da sequência didática, os alunos foram novamente submetidos ao TEPEP, agora como Pós-Teste. No Gráfico 21, está a comparação entre os resultados do Pré-Teste e Pós-Teste.

Gráfico 21 – Resultados do Pré-teste e Pós-teste.

Fonte: Próprio do autor.

Ao comparar os resultados, pode-se notar que dos 17 alunos, 11 evoluíram, um (1) permaneceu com o mesmo escore e 5 alunos (A06, A09, A11, A15 e A17) obtiveram maior escore no Pré-Teste que no Pós-Teste. A média no Pré-Teste foi de 8,23 e no Pós-

Teste 9,29. Considerando os alunos que pontuaram bem na primeira vez e mau na segunda, os únicos com escore acima da média no Pré-Teste e abaixo da média no Pós-Teste foram o A09 e A15, todos os outros, já faziam parte dos alunos com notas abaixo da média no Pré-Teste. Os alunos com escore muito baixo no teste que passaram a atingir a média foram: A04, A14 e A16. Os alunos que mostraram uma evolução surpreendente foram: A04, A13, A14 e A16. Conforme afirmam Goodsell et al. (1992), na aprendizagem ativa, os alunos são diversos: aprendem de forma diferente, têm concepções diferentes, possuem experiências diversas, etc.; assim, não se esperava que todos os alunos evoluíssem da mesma forma. Em relação ao Pré-Teste, 9 alunos tiveram notas abaixo da média; no Pós-Teste, 8 alunos ficaram abaixo da média. Considerando as médias no Pré-Teste e Pós-Teste, a turma como um todo, teve uma taxa de evolução de 13 %. No Gráfico 22, são exibidos os escores nos itens do Pré-Teste e Pós-Teste.

Gráfico 22 – Escore dos itens no Pré-teste e Pós-teste.

Fonte: Próprio do autor.

Dos 15 itens do TEPEP, nove apresentaram aumento no escore, quatro permaneceram com o mesmo escore e dois (I04 e I12) tiveram escore maior no Pré-Teste do que no Pós-Teste. O escore médio dos itens no Pré-Teste foi de 9,33, no Pós-Teste 10,53. Em relação ao primeiro escore médio, oito itens mostram um escore abaixo da média; considerando o segundo escore médio, do Pós-Teste, sete itens exibem um escore

abaixo da média. Dos itens com escore maior no Pré-Teste e menor no Pós-Teste, apenas o item I04 estava, anteriormente, acima da média. Os itens que demonstraram um aumento elevado no escore foram: I02, I05, I9 e I14. Levando em conta os escores médios, a taxa de evolução nos escores dos itens foi de 13 %.

6 CONSIDERAÇÕES FINAIS

O avanço tecnológico fez com que o uso da Estatística e Probabilidade ganhasse grandes proporções. Hoje, através de um *smartphone* ou *tablet*, tem-se acesso instantâneo às informações estatísticas. Não por acaso, a preocupação com o ensino desse tema tem crescido exponencialmente. Mas infelizmente, o ensino da Estatística e Probabilidade nas escolas de ensino público do Brasil, em geral, é deixado de lado, para última instância. Um grande erro, pois vive-se na era em que as informações valem muito e transitam em segundos. Por isso, o ensino da Estatística e Probabilidade exige o emprego de estratégias inovadoras, cativantes.

Este trabalho teve por objetivo propor um projeto com uma sequência didática atrativa aos alunos, capaz de desenvolver suas habilidades e proporcionar o aprendizado de conceitos da Estatística e Probabilidade. Tal proposta foi direcionada a alunos das séries finais do Ensino Fundamental II e teve respaldo na aprendizagem ativa/colaborativa, com base em Planejamento de Experimentos. Para atingir o objetivo do trabalho, buscou-se responder à seguinte questão: o uso de Planejamento de Experimentos pode contribuir para o desenvolvimento das habilidades dos alunos e proporcionar a aprendizagem ativa de conceitos básicos da Estatística e Probabilidade?

Buscando responder à questão, indagou-se: qual atividade experimental seria interessante aos alunos? Inspirado no experimento de George Box, com helicópteros de papel, pensou-se em proporcionar uma atividade parecida, voltada para o contexto deles: um projeto com aviões de papel. Os alunos tiveram de selecionar um único modelo de avião de papel, com o objetivo de prolongar seu tempo de voo. Ao fim de todas as atividades do projeto, foi verificado se houve aumento do nível de proficiência em Estatística e Probabilidade dos alunos, assim como o exercício de suas habilidades “extracurriculares”. Para medir essa proficiência, foi desenvolvido e validado um teste: Teste de Proficiência em Estatística e Probabilidade (TEPEP), também objeto de investigação neste trabalho.

Em relação ao TEPEP, pôde-se concluir que o teste atendeu às características de uma avaliação diagnóstica. É um instrumento confiável, que mede apenas o construto Proficiência em Estatística e Probabilidade. Porém, um de seus itens (01) deve ser reformulado ou retirado do teste, pois seus índices de dificuldade e discriminação são

baixíssimos, para os sujeitos da pesquisa. A despeito da remoção ou permanência da questão, o teste é uma boa ferramenta para se medir o conhecimento estatístico e probabilístico de alunos das séries finais do Ensino Fundamental II. A maioria das avaliações externas apresenta poucos itens sobre o tema, sendo insuficientes para uma avaliação ampla. Uma possibilidade para o aprofundamento desse trabalho, quanto ao teste, seria de sua validação para alunos de Ensino Médio, ou até mesmo para alunos ingressantes em cursos de graduação.

Sobre os métodos psicométricos empregados para a validação do TEPEP, conclui-se que são ferramentas com capacidade extraordinária. Apesar de estarem ao alcance de todos e serem utilizados em avaliações como o Exame Nacional do Ensino Médio (ENEM) e Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), ainda são pouco conhecidos pelos professores. Esses métodos deveriam ser adotados pelas redes municipais de educação, como forma de desenvolver avaliações independentes. Em relação à Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI), notou-se que são complementares, sendo a TRI mais precisa em seus resultados. Seria precipitado dizer que, nesse momento, a TRI irá extinguir a TCT. Um ponto de partida para outras pesquisas é a investigação da utilização do método de rotação Varimin, ainda pouco explorado, que demonstra ter potencial.

Os resultados do desempenho dos alunos, no Pré-Teste e Pós-Teste, mostraram que a metodologia utilizada nesta pesquisa teve significância estatística sobre o aumento do nível de proficiência em Estatística e Probabilidade dos 17 alunos que participaram do projeto. Além da sequência didática do projeto ter proporcionado aprendizado de conceitos estatísticos e probabilísticos, descritos nos objetivos específicos dessa dissertação, possibilitou também o desenvolvimento de habilidades como liderança, trabalho em equipe, organização, argumentação, etc.. Ou seja, o uso de Planejamento de Experimentos, para o ensino ativo da Estatística e Probabilidade, apresentou bons resultados. É um método que proporciona a criação e discussão de hipóteses, promovendo um ambiente rico para o ensino de qualquer ciência. Além disso, em seu processo podem surgir novas situações e conjecturas, abrindo um ciclo para o processo de ensino e aprendizagem.

Como existem inúmeras possibilidades de experimentos para serem realizados através de Planejamento de Experimentos, os professores que a adotarem terão recursos inesgotáveis para se trabalhar com Estatística e Probabilidade. Compete à criatividade e

capacidade de pesquisa de cada docente criar ou conhecer outros experimentos. Para uma primeira iniciativa, sugere-se o uso de Planejamento de Experimentos completos.

Entende-se que este trabalho procurou dar uma nova visão sobre como os conceitos estatísticos e probabilísticos podem ser ensinados e avaliados. Além disso, as práticas desenvolvidas com os alunos procuraram desenvolver não só conteúdos, mas habilidades que vão para fora de um currículo puramente acadêmico. Contribuiu para a sociedade ao formar cidadãos mais capazes de tomar decisões fundamentadas, respeitar a opinião dos outros, serem passivos no momento de ouvir e ativos no momento de argumentar. Para a ciência em questão – ensino de Probabilidade e Estatística – traz um novo modelo didático, disponível a todos que desejarem. Espera-se que, através deste texto, professores sintam-se desafiados a sair da “zona de conforto”, que se incomodem com suas práticas e pesquisem ou criem métodos diferentes, que atendam melhor seus alunos e possam contribuir para uma educação de maior qualidade.

Sobre a Estatística e Probabilidade, compreende-se que não deva ser vista como um ramo da Matemática, esquecida para último caso. Pelo contrário, deve ser concebida como protagonista e independente, visto que é suficientemente dotada de bases empíricas e teóricas.

Enfim, considera-se que os objetivos desse trabalho foram atingidos. Conclui-se que o uso de Planejamento de Experimentos pôde proporcionar aos alunos um ensino ativo e colaborativo dos conceitos da Estatística e Probabilidade.

7 PRODUTOS DECORRENTES DESTE TRABALHO

- Metodologia para o ensino da Estatística e Probabilidade, baseada em planejamento de experimentos;
- Metodologia para a validação de instrumentos de medida, baseada na Psicometria;
- Teste de Estatística e Probabilidade já validado.

8 REFERÊNCIAS

BAKER, F. B.. **The Basics of Item Response Theory**. Washington, Dc: Eric, 2001.

BARROWS, H. S.; TAMBLYN, R. M.. **Problem-based learning: An approach to medical education**. New York: Springer Publishing Company, 1980.

BISGAARD, S.. Teaching Statistics to Engineers. **The American Statistician**, v. 45, n. 4, p. 274-283, 1991.

BARTLETT, M. S.. A Note on the Multiplying Factors for Various χ^2 Approximations. **Journal of The Royal Statistical Society**, v. 16, n. 2, p. 296-298, 1954.

_____. Tests of Significance In Factor Analysis. **British Journal of Statistical Psychology**, v. 3, n. 2, p. 77-85, 1950.

BOCK, R. D.. A Brief History of Item Theory Response. **Educational Measurement: Issues and Practice**, v. 16, n. 4, p. 21-33, 2005.

BOER, F. G.; CATEN, C. T.; PAULA, I. C.. Application of the Problem Based Learning Method in the Discipline 'Statistics for Engineering'. In: ASEE 2014 ZONE I CONFERENCE, 1., 2014, Bridgeport. **Proceedings...** Bridgeport: University of Bridgeport, 2014. p. 1 - 4.

BONWELL, C. C.; EISON, J. A.. **Active Learning: Creating Excitement in Classroom**. Washington, D.C: Ashe-eric Higher Education Report, 1991.

BOX, G. E. P.. George's Column. **Quality Engineering**, v. 4, n.3, p. 453-459, 1992.

BRASIL. Secretaria de Educação Fundamental. **Parâmetros Curriculares Nacionais (5ª a 8ª séries): Matemática**. Brasília: MEC, 1998.

BROWN, T. A.. **Confirmatory Factor Analysis for Applied Research**. New York: The Guilford Press, 2015.

CAMPBELL, D. T.; STANLEY, J. C.. **Experimental and Quasi-Experimental Designs for Research**. Boston: Houghton Mifflin Company, 1963.

CASSEPP-BORGES, V., BALBINOTTI, M.A.A.; TEODORO, M.L.M.. Tradução e validação de conteúdo: uma proposta para adaptação de instrumentos. In: Pasquali, L.

(Org.), **Instrumentação psicológica. Fundamentos e práticas**. Porto Alegre: Artmed, 2010. p. 506-520.

CATTELL, R. B.. The Scree Test For The Number of Factors. **Multivariate Behavioral Research**, v. 1, n. 2, p. 245-276, 1966.

CHACÓN, I. M. G.. Cuestiones Afectivas en la Enseñanza de las Matemáticas: Una perspectiva para el profesor. In: Extremadura, Universidad de. En L. C. Contreras y L. J. Blanco, **Aportaciones a la formación inicial de maestros en el área de matemáticas: Una mirada a la práctica docente**. Cáceres: Universidad de Extremadura, 2002. p. 23-58.

COBB, G.. Teaching Statistics. In: Steen, L. A.. **Heeding the Call for Change: Suggestions for Curricular Action**, Washington, D. C.: Mathematical Association of America, 1992. cap. 1, p. 3-43.

_____; MOORE, D. S.. Mathematics, Statistics, and Teaching. **The American Mathematical Monthly**, v. 104, n. 9, p. 801-823, 1997.

COHEN, J.. A Power of Prime. **Psychological Bulletin**, Washington, v. 112, n. 1, p. 155-159, 1992.

COHEN, L.; MANION, L.; MORRISON, K.. **Research Methods in Education**. New York: Taylor & Francis E-library, 2007.

COMREY, A. L.. **A First Course in Factor Analysis**. New York: Academic Press, 1973.

CRESWELL, J. W.. **Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research**. Lincoln: Pearson, 2012.

CRONBACH, L. J.; MEEHL, P. E.. Construct validity in psychological tests. **Psychological Bulletin**, v. 52, n. 4, p. 281-302, 1955.

CROSS, K. P.. **Teaching "For" Learning**. Raleigh: The Educational Resources Information Center (ERIC), 1987.

DEWEY, J.. **Democracia e educação**. Tradução Godofredo Rangel e Anísio Teixeira. São Paulo: Nacional, 1959.

EBEL, R. L.; FRISBIE, D. A.. **Essentials of Educational Measurement**. New Jersey: Prentice Hall College, 1991.

FABRIGAR, L. R. et al. Evaluating the use of exploratory factor analysis in psychological research. **Psychological Methods**, v. 4, n. 3, p. 272-299, 1999.

FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. da. Visão além do alcance: uma introdução à análise fatorial. **Opin. Publica**, v. 16, n. 1, p. 160-185, jun. 2010.

FISHER, R. A.; YATES, F.. **Statistical Tables:** For Biological, Agricultural and Medical research. Londres: Oliver And Boyd, 1963.

FLOYD, F. J.; WIDAMAN, K. F.. Factor analysis in the development and refinement of clinical assessment instruments. **Psychological Assessment**, v. 7, n. 3, p. 286-299, 1995.

FOLD'NFLY. **Paper Airplane Designs:** A database of paper airplanes with easy to follow folding instructions, 2015. Disponível em: <<http://www.foldnfly.com/index.html#/1-1-1-1-1-1-1-2>>. Acesso em: 17 jul. 2016.

GAL, I.; GARFIELD, J.. Curricular Goals and Assessment Challenges in Statistics Education. In: GAL, I.; GARFIELD, J. (Ed.). **The Assessment Challenge in Statistics Education**. Minnesota: IOS Press, 1997. p. 1-13.

GARFIELD, J.; AHLGREN, A.. Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research. **Journal for Research In Mathematics Education**, v. 19, n. 1, p. 44-63, 1988.

GARFIELD, J.; BEN-ZVI, D.. How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics. **International Statistical Review**, v. 75, n. 3, p. 372-396, 2007.

GOODSELL, A. S. et al.. **Collaborative Learning:** A Sourcebook for Higher Education. Washington, D.C: ERIC, 1992.

GUADAGNOLI, E.; VELICER, W. E.. Relation of Sample Size to the Stability of Component Patterns. **Psychological Bulletin**, v. 103, n. 2, p. 265-275, 1988.

GULLIKSEN, H.. **Theory of mental tests**. New York: John Wiley & Sons, 1950.

HAIR, J. F. et al. **Multivariate Data Analysis**. New Jersey: Pearson Prentice Hall, 2009.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J.. **Fundamentals of Item Response Theory**. Londres: Sage Publications, 1991.

HAMBLETON, R. K.; JONES, R. W.. An NCME Instructional Module on. **Educational Measurement: Issues and Practice**, v. 12, n. 3, p. 38-47, 2005.

HAMBLETON, R. K.; LINDEN, W. J. V D.. Advances in Item Response Theory and Applications: An Introduction. **Applied Psychological Measurement**, v. 6, n. 4, p. 373-378, 1982.

HENSON, R. K.. Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. **Educational And Psychological Measurement**, v. 66, n. 3, p. 393-416, 2006.

HERNÁNDEZ-NIETO, R. A.. **Contributions to Statistical Analysis**. Mérida: Universidade de Los Andes, 2001.

HINKLE, D. E.; WIERSMA, W.; JURIS, S. G. **Applied statistics for the behavioral sciences**. Boston: Houghton Mifflin, 1998.

HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. M.. **Psicometria**. Porto Alegre: Artmed, 2015.

ISAAC, S.; MICHAEL, W. B.. **Handbook in Research and Evaluation**: For education and the Behavioral Sciences. San Diego: Edits Publishers, 1995.

JOHNSON, D. W.; JOHNSON, R. T.; SMITH, K. A.. Cooperative Learning Returns To College What Evidence Is There That It Works? **Change: The Magazine of Higher Learning**, v. 30, n. 4, p. 26-35, 1998.

KAISER, H. F.. A second generation little jiffy. **Psychometrika**, v. 35, n. 4, p. 401-415, 1970.

_____. An Index of Factorial Simplicity. **Psychometrika**, v. 39, n. 1, p. 31–36, 1974.

_____. The Application of Electronic Computers to Factor Analysis. **Educational And Psychological Measurement**, v. 20, n. 1, p. 141-151, 1960.

KELLEY, T. L.. The selection of upper and lower groups for the validation of test items. **Journal of Educational Psychology**, v. 30, n. 1, p. 17-24, 1939.

KERLINGER, F. N.; LEE, H. B.. **Foundations of Behavioral Research**. Nova Iorque: Holt, Rinehart And Wiston, 1973.

KNOKE D, B. G. W.; MEE, A.P.. **Statistics for Social Data Analysis**. Itasca, IL: F.E. Peacock Publishers. 2002.

KRISHNAN, V.. **The Early Child Development Instrument (EDI):** An item analysis using Classical Test Theory (CTT) on Alberta's data. Edmonton: Early Child Development Mapping (ECMap) Project, 2013.

LAZARSELD, P. F.. The logical and mathematical foundation of latent structure analysis. In: STOUFFER, S. A. et al. **Measurement And Prediction**. Princeton: Princeton University Press, 1950, p. 207-209.

LEE, K.; MITCHELL, D.. **Paper Airplane Fold-a-Day 2014 Day-to-Day Calendar:** Calendar – Day to Day Calendar, July 16, 2013. Kansas City: Accord Publishing, 2013.

LINACRE, J. M.. **A User's Guide to WINSTEPS MINISTEP:** Rasch-Model Computer Programs. Chicago: Winsteps.com, 2012.

LUDLOW, D. K.; SCHULZ, K. H.; ERJAVEC, J.. Teaching Statistical Experimental Design Using a Laboratory Experiment. **Journal of Engineering Education**, v. 84, n. 4, p. 351-359, 1995.

LORD, F. M.. **A Theory of Test Scores (Psychometric Monograph No. 7)**. Richmond: Psychometric Corporation, 1952.

_____. Estimation of latent ability and item parameters when there are omitted responses. **Psychometrika**, v. 39, n. 2, p. 247-264, 1974.

MACDONALD, P.; PAUNONEN, S. V.. A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory. **Educational And Psychological Measurement**, v. 62, n. 6, p. 921-943, 2002.

MONTGOMERY, D. C.. **Design and Analysis of Experiments**. New York: John Wiley & Sons, Inc, 2001.

NARDI, A. H.; WALES, C. E.. Teaching Decision Making with Guided Design. In: COSTA, A. L. (Ed.). **Developing Minds: A Resource Book for Teaching Thinking**. Alexandria, Va: Association for Supervision and Curriculum development, 1985. p. 222-225.

NIEMI, H.. Active learning – a cultural change needed in teacher education and schools. **Teaching And Teacher Education**, v. 18, n. 7, p. 763-780, 2002.

NOVICK, M. R.. The Axioms And Principal Results of Classical Test Theory. **Ets Research Bulletin Series**, v. 1965, n. 1, p. 1-31, 1965.

O'ROURKE, N. et al. A Cross-National Comparison of Reminiscence Functions Between Canadian and Israeli Older Adults. **The Journals of Gerontology Series B: Psychological Sciences and Social Sciences**, v. 68, n. 2, p. 184-192, 2012.

OSBORNE, J. W.; COSTELLO, A. B.. Best Practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. **Pan-pacific Management Review**, Raleigh, v. 2, n. 12, p. 131-146, 2009.

PATHAK, R. P.. **Methodology of Educational Research**. New Delhi: Atlantic, 2008.

PASQUALI, L.. **Psicometria: Teoria dos Testes na Psicologia e na Educação**. Petrópolis: Vozes, 2003.

_____. Psicometria. **Revista da Escola de Enfermagem da Usp**, v. 43, p. 992-999, 2009.

_____; PRIMI, R.. Fundamentos da Teoria da Resposta ao Item –TRI. **Avaliação Psicológica**, Porto Alegre, v. 2, n. 2, p. 99-110, 2003.

PENZIAS, A.. Teaching Statistics to Engineers. **Science**, v. 244, n. 4908, p. 1025-1025, 1989.

PETT, M. A.; LACKEY, N. R.; SULLIVAN, J. J.. **Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research**. California: Sage-usa, 2003. 328 p..

PRINCE, M.. Does Active Learning Work? A Review of the Research. **Journal of Engineering Education**, v. 93, n. 3, p. 223-231, 2004.

RAYMUNDO, V. P.. Construção e validação de instrumentos: um desafio para a psicolinguística. **Letras de Hoje**, Porto Alegre, v. 44, n. 3, p. 86-93, 2009.

RODGERS, J. L.; NICEWANDER, W. A.. Thirteen Ways to Look at the Correlation Coefficient. **The American Statistician**, v. 42, n. 1, p. 59-66, 1988.

ROSNER, B.. **Fundamentals of Biostatistics**. Belmont: Thomson, 1995.

SANZ, J. A.. **La activación del aprendizaje como estrategia efectiva para un adecuado ejercicio docente**. Caracas, VE: Universidad Simón Bolívar (U.S.B), 1999.

SÃO PAULO (Estado). Secretaria de Educação do Estado de São Paulo. **Relatório Pedagógico: 2011 SARESP**. São Paulo: 2011.

_____. Secretaria de Educação do Estado de São Paulo. **Relatório Pedagógico: 2012 SARESP**. São Paulo: 2012.

_____. Secretaria de Educação do Estado de São Paulo. **Relatório Pedagógico: 2013 SARESP**. São Paulo: 2013.

_____. Secretaria de Educação do Estado de São Paulo. **Relatório Pedagógico: 2014 SARESP**. São Paulo: 2014.

SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O.. Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. **Psicol. Reflex. Crit.**, v. 26, n. 2, p. 241-250, 2013.

SILBERMAN, M.. **Aprendizaje Activo: 101 estrategias para enseñar cualquier tema**. Tradução Adriana Oklander. Capital Federal, Argentina: Troquel, 1998.

SMITH, B. L.; MACGREGOR, J.. What is Collaborative Learning? In: GOODSELL, Anne et al. **Collaborative Learning: A Sourcebook for Higher Education**. Washington: Eric, 1992. p. 9-22.

SPIEGEL, M. R.. **Probabilidade e Estatística**. São Paulo: Mcgraw-hill, 1978.

STERN, D.; HUBER, G. L.. **Active Learning for Students an Teachers**. Paris: OCDE, 1997.

STEYER, R.. Classical (psychometric) test theory. In: RAGIN, C.; COOK, T. (Ed.). **International encyclopedia of the social and behavioral sciences. Logic of inquiry and research design**. Oxford, England: Pergamon, 2001. p. 1955-1962

TABACHNICK, B. G.; FIDELL, L. S.. **Using multivariate statistics**. Boston: Pearson, 2013.

THOMPSON, Bruce. **Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications**. Washington, DC: American Psychological Association, 2004.

THURSTONE, L. L.. A Method of Scaling Psychological and Educational Tests. **Journal of Educational Psychology**. Chicago, p. 433-451, 1925.

TOLEDO, G. L.; OVALLE, I. I.. **Estatística Básica**. São Paulo: Atlas, 1983.

URBINA, S.. **Fundamentos da Testagem Psicológica**. Porto Alegre: Artmed, 2007

VIANNA, H. M.. **Testes em Educação**. São Paulo: IBRASA, 1978.

_____. Validade de construto em testes educacionais. **Educação e Seleção**, n. 8, p. 35-44, 1983.

VELICER, W. F.; FAVA, J. L.. Affects of variable and subject sampling on factor pattern recovery. **Psychological Methods**, v. 3, n. 2, p. 231-251, 1998.

WALICHINSKI, D.; SANTOS JUNIOR, G.. Educação Estatística: objetivos, perspectivas e dificuldades. **Imagens da Educação**, v. 3, n. 3, p. 31-37, 2013.

WILLIAMS, B.; BROWN, T.; ONSMAN, A.. Exploratory factor analysis: A five-step guide for novices. **Australasian Journal of Paramedicine**, v. 8, n. 3, 2010.

APÊNDICE A – Teste de Proficiência em Estatística e Probabilidade (TEPEP)

Orientações sobre o teste:

- 1- Não há necessidade de se identificar;
- 2- Cada questão é composta por seu enunciado seguido de quatro alternativas de respostas, estando somente **uma** delas correta;
- 3- Na folha de respostas do teste (**última página**) assinale com um **X** apenas a **única** alternativa que considera correta.
- 4- Os espaços em branco **podem** ser utilizados para realização de cálculos.

QUESTÕES

- 1) A tabela indica os códigos de Discagem à Distância (DDD) de algumas cidades do Estado de São Paulo.

Localidades	DDD	Localidades	DDD
Araçatuba	18	Santos	13
Araraquara	16	São José do Rio Preto	17
Campinas	19	São José dos Campos	12
Cubatão	13	São Paulo	11

Dessas cidades, as que possuem os mesmos DDD são:

- (A) Campinas e São Paulo.
 (B) Cubatão e Santos.
 (C) Cubatão e São José do Rio Preto.
 (D) Araraquara e Santos.

- 2) Um estacionamento tem capacidade para 180 veículos. No momento, 50 % das vagas estão ocupadas. O número de vagas ocupadas é:

- (A) 90
 (B) 95
 (C) 130
 (D) 135

- 3) Com as promoções que muitas companhias aéreas têm feito, fica cada vez mais fácil viajar de avião. Observe na figura abaixo o número de passageiros nos últimos anos.

A tabela que melhor representa a figura é:

(A)

Ano	Meses	Nº de Passageiros (em milhões)
2007	janeiro a dezembro	71
2008	janeiro a dezembro	128
2009	janeiro a dezembro	113
2010	janeiro a julho	110

(B)

Ano	Meses	Nº de Passageiros (em milhões)
2007	janeiro a dezembro	113
2008	janeiro a dezembro	110
2009	janeiro a dezembro	71
2010	janeiro a julho	128

(C)

Ano	Meses	Nº de Passageiros (em milhões)
2007	janeiro a dezembro	110
2008	janeiro a dezembro	113
2009	janeiro a dezembro	71
2010	janeiro a julho	128

(D)

Ano	Meses	Nº de Passageiros (em milhões)
2007	janeiro a dezembro	110
2008	janeiro a dezembro	113
2009	janeiro a dezembro	128
2010	janeiro a julho	71

- 4) As cartas abaixo serão colocadas numa caixa e uma será retirada ao acaso.

A probabilidade de a carta retirada ter a figura de uma pessoa é:

- (A) $\frac{2}{6}$ (B) $\frac{2}{8}$ (C) $\frac{4}{10}$ (D) $\frac{4}{6}$

- 5) Na tabela a seguir está a relação de gols marcados pelo maior artilheiro da Copa do mundo de Futebol, o alemão Miroslav Klose.

GOLS EM COPAS - MIROSLAV KLOSE	
Edição da Copa	Nº de gols
Coréia/Japão (2002)	5
Alemanha (2006)	5
África do Sul (2010)	4
Brasil (2014)	2

Pode-se afirmar, com base nos dados da tabela, que a média de gols do atacante em copas do mundo é de:

- (A) 4 gols.
 (B) 6 gols.
 (C) 5 gols.
 (D) 3 gols.
- 6) Uma máquina fotográfica custava R\$ 500,00. No dia dos pais, numa promoção, foi vendida com um desconto de 10 % e, logo depois, em cima do novo preço sofreu um aumento de 10 %.

O seu preço atual, em reais, é:

- (A) 450,00
 (B) 475,00
 (C) 495,00
 (D) 515,00
- 7) Lúcia precisava descobrir quantos números de dois algarismos diferentes podem ser formados utilizando apenas os algarismos 3, 5, 7 e 8. Ela resolveu, então, representar um diagrama de árvore para facilitar a contagem. Lúcia iniciou assim:

Depois de completar o diagrama, a quantidade de números de dois algarismos distintos que Lúcia encontrou foi:

- (A) 8
 (B) 12
 (C) 10
 (D) 14

- 8) O número de passageiros que utilizam ônibus municipal na Capital e na Grande São Paulo diminuiu nos últimos cinco anos, como mostram os gráficos a seguir:

Passageiros por ano em 2006

Passageiros por ano em 2011

(Folha S. Paulo, Cotidiano C3, 02.05.2012 adaptado)

A respeito das informações do texto e dos gráficos, é correto afirmar que a quantidade de passageiros de

- (A) metrô aumentou; trem aumentou; ônibus metropolitano não foi alterada e de ônibus municipal não foi alterada.
 (B) metrô aumentou; trem aumentou; ônibus metropolitano diminuiu e de ônibus municipal diminuiu.
 (C) metrô aumentou; trem não foi alterada; ônibus metropolitano não foi alterada e de ônibus municipal diminuiu.
 (D) metrô aumentou; trem aumentou; ônibus metropolitano não foi alterada e de ônibus municipal diminuiu.

- 9) Um dado de 6 faces foi lançado 20 vezes, obtendo-se os seguintes resultados na face superior:

3	4	6	2	3	5	2	4	1	5	2	6	3	4	5	1	1	1	5	1

A frequência relativa do resultado 5 é:

- (A) $\frac{4}{20}$
 (B) $\frac{4}{6}$
 (C) $\frac{1}{4}$
 (D) $\frac{5}{20}$

- 10)** Um estojo de maquiagem tem 12 tonalidades de batom, sendo 3 tonalidades cintilantes e as restantes cremosas.

A probabilidade de se retirar, ao acaso, desse estojo um batom cintilante é:

- (A) 30 %
 (B) 10 %
 (C) 25 %
 (D) 20 %

- 11)** Durante 15 dias, uma concessionária de automóveis registrou quantos clientes estiveram na loja:

dias	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º	11º	12º	13º	14º	15º
clientes	8	4	6	7	6	3	6	2	7	6	5	5	6	9	7

A tabela de frequência que representa corretamente os dados registrados pela concessionária é:

(A)

Clientes	Frequência
0 a 2	1
3 a 4	1
5 a 6	7
7 a 8	5
9 a 10	1
TOTAL	15

(B)

Clientes	Frequência
0 a 2	1
3 a 4	2
5 a 6	7
7 a 8	4
9 a 10	1
TOTAL	15

(C)

Clientes	Frequência
0 a 2	1
3 a 4	2
5 a 6	6
7 a 8	4
9 a 10	2
TOTAL	15

(D)

Clientes	Frequência
0 a 2	2
3 a 4	2
5 a 6	7
7 a 8	4
9 a 10	1
TOTAL	15

- 12) Luísa foi à sorveteria. Lá havia três sabores de sorvete: chocolate, morango e flocos; e dois tipos de cobertura: caramelo e chocolate.

O número de maneiras diferentes de Luísa escolher o seu sorvete com apenas um sabor e um tipo de cobertura é:

- (A) 6
(B) 7
(C) 8
(D) 4

- 13) A mãe de Ana foi anotando a altura de sua filha durante o primeiro ano de vida. Veja a tabela.

Idade	Altura
Ao nascer	49 cm
1 mês	52 cm
3 meses	56 cm
5 meses	62 cm
7 meses	66 cm
9 meses	69 cm

Entre os gráficos abaixo, aquele que melhor apresenta as informações da tabela é:

14) A média aritmética entre 14, 12, 16, 18 e 20 é:

- (A) 14
- (B) 15
- (C) 12
- (D) 16

15) Na tabela a seguir, estão as previsões de temperatura mínima e máxima para alguns dias do mês de março, na cidade de Lorena, Estado de São Paulo.

Previsão de Temperatura - Lorena/SP		
Dias	Mínima	Máxima
Terça-feira-08/03/16	19°C	29°C
Quarta-feira-09/03/16	19°C	30°C
Quinta-feira-10/03/16	20°C	30°C
Sexta-feira-11/03/16	21°C	28°C
Sábado-12/03/16	18°C	28°C
Domingo-13/03/16	17°C	29°C

Com base nos dados, pode-se afirmar que a moda da temperatura mínima e a mediana da temperatura máxima são respectivamente:

- (A) 20°C e 28°C.
- (B) 19°C e 29°C.
- (C) 20°C e 29°C.
- (D) 19°C e 30°C.

FOLHA DE RESPOSTAS

	A	B	C	D
01	☐	☐	☐	☐
	A	B	C	D
02	☐	☐	☐	☐
	A	B	C	D
03	☐	☐	☐	☐
	A	B	C	D
04	☐	☐	☐	☐
	A	B	C	D
05	☐	☐	☐	☐
	A	B	C	D
06	☐	☐	☐	☐
	A	B	C	D
07	☐	☐	☐	☐
	A	B	C	D
08	☐	☐	☐	☐
	A	B	C	D
09	☐	☐	☐	☐
	A	B	C	D
10	☐	☐	☐	☐
	A	B	C	D
11	☐	☐	☐	☐
	A	B	C	D
12	☐	☐	☐	☐
	A	B	C	D
13	☐	☐	☐	☐
	A	B	C	D
14	☐	☐	☐	☐
	A	B	C	D
15	☐	☐	☐	☐

APÊNDICE B – Tabela para registro do tempo de voo dos aviões de papel

EXPERIMENTO Nº: _____																	
REGISTRO DO TEMPO DE VOO																	
	AVIÃO	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
RODADAS	Cronômetros	TEMPO EM SEGUNDOS (s)															
1ª Rodada	C ₁																
	C ₂																
	C ₃																
	$\overline{C_{R_1}}$																
2ª Rodada	C ₁																
	C ₂																
	C ₃																
	$\overline{C_{R_2}}$																
3ª Rodada	C ₁																
	C ₂																
	C ₃																
	$\overline{C_{R_3}}$																
4ª Rodada	C ₁																
	C ₂																
	C ₃																
	$\overline{C_{R_4}}$																
TEMPO MÉDIO FINAL																	

C_n = Cronômetros; $\overline{C_{R_n}}$ = média dos cronômetros.

APÊNDICE C – Roteiro para organização dos grupos durante o experimento

Grupo:			
Líder do grupo:			
Demais integrantes:			
Quadro de distribuição das tarefas			
Avião	Lançar	Cronometrar	Calcular a média
01	Grupo_____	Grupo_____	Grupo_____
02	Grupo_____	Grupo_____	Grupo_____
03	Grupo_____	Grupo_____	Grupo_____
04	Grupo_____	Grupo_____	Grupo_____
05	Grupo_____	Grupo_____	Grupo_____
06	Grupo_____	Grupo_____	Grupo_____
07	Grupo_____	Grupo_____	Grupo_____
08	Grupo_____	Grupo_____	Grupo_____
09	Grupo_____	Grupo_____	Grupo_____
10	Grupo_____	Grupo_____	Grupo_____
11	Grupo_____	Grupo_____	Grupo_____
12	Grupo_____	Grupo_____	Grupo_____
13	Grupo_____	Grupo_____	Grupo_____
14	Grupo_____	Grupo_____	Grupo_____
15	Grupo_____	Grupo_____	Grupo_____
16	Grupo_____	Grupo_____	Grupo_____

APÊNDICE D – Dimensões e linhas de corte (ARUMAKAN)

APÊNDICE E – Saída do Programa Minitab (Plano do Experimento)

18/04/2016 07:10:02

Multilevel Factorial Design

Factors: 4 Replicates: 1
 Base runs: 16 Total runs: 16
 Base blocks: 1 Total blocks: 1

Number of levels: 2; 2; 2; 2

Design Table (randomized)

Run	A	B	C	D
1	2	1	2	1
2	1	1	1	2
3	1	2	2	1
4	1	2	1	1
5	1	2	2	2
6	2	1	2	2
7	2	1	1	1
8	2	2	2	2
9	2	1	1	2
10	2	2	1	2
11	1	2	1	2
12	2	2	2	1
13	1	1	1	1
14	2	2	1	1
15	1	1	2	1
16	1	1	2	2

Numeração	Lançamento	Asa Dobrada	Tamanho	Corpo Fixo	Tipo de Papel
11	1º	sim	original	sim	Sulfite
02	2º	não	original	não	novo antílope
07	3º	não	reduzido	sim	Sulfite
05	4º	não	reduzido	não	Sulfite
08	5º	não	reduzido	sim	novo antílope
12	6º	sim	original	sim	novo antílope
09	7º	sim	original	não	Sulfite
16	8º	sim	reduzido	sim	novo antílope
10	9º	sim	original	não	novo antílope
14	10º	sim	reduzido	não	novo antílope
06	11º	não	reduzido	não	novo antílope
15	12º	sim	reduzido	sim	Sulfite
01	13º	não	original	não	Sulfite
13	14º	sim	reduzido	não	Sulfite
03	15º	não	original	sim	Sulfite
04	16º	não	original	sim	novo antílope

APÊNDICE F – Resultados do julgamento dos juízes

	Juiz 1			Juiz 2			Juiz 3			Juiz 4			Juiz 5			Juiz 6		
Item	CL	PP	RT	CL	PP	RT	CL	PP	RT	CL	PP	RT	CL	PP	RT	CL	PP	RT
1	5	5	5	5	4	4	5	5	5	5	4	3	5	5	5	4	4	4
2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4
3	5	5	5	3	4	5	4	5	5	5	5	5	5	5	5	4	4	4
4	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	4	5
5	5	5	5	4	4	4	5	5	5	5	5	5	5	5	5	5	4	4
6	5	5	5	3	4	4	3	5	5	4	4	5	5	5	5	4	4	5
7	5	5	5	3	4	4	5	5	5	4	5	5	4	5	5	4	4	5
8	5	5	5	2	4	4	4	3	5	5	5	5	5	5	5	4	5	5
9	5	5	5	2	3	3	5	5	5	5	5	5	5	5	5	5	5	5
10	5	5	5	2	3	3	5	5	5	4	4	4	5	5	5	5	4	4
11	5	5	5	3	4	4	4	5	5	5	4	4	5	5	5	5	4	5
12	5	5	5	3	4	4	5	5	5	4	4	4	5	5	5	5	4	5
13	5	5	5	4	5	5	4	5	5	5	5	5	5	5	5	5	5	4
14	5	5	5	4	5	5	5	5	5	4	4	4	5	5	5	5	5	5
15	5	5	5	3	4	4	4	5	5	5	5	5	5	5	5	5	4	5

Média das notas dos juízes nos itens						
Item	Juiz 1	Juiz 2	Juiz 3	Juiz 4	Juiz 5	Juiz 6
1	5,0	4,3	5,0	3,0	5,0	4,0
2	5,0	5,0	5,0	5,0	5,0	4,3
3	5,0	4,0	4,7	5,0	5,0	4,0
4	5,0	4,7	5,0	5,0	5,0	4,7
5	5,0	4,0	5,0	5,0	5,0	4,3
6	5,0	3,7	4,3	5,0	5,0	4,3
7	5,0	3,7	5,0	5,0	4,7	4,3
8	5,0	3,3	4,0	5,0	5,0	4,7
9	5,0	2,7	5,0	5,0	5,0	5,0
10	5,0	2,7	5,0	4,0	5,0	4,3
11	5,0	3,7	4,7	4,0	5,0	4,7
12	5,0	3,7	5,0	4,0	5,0	4,7
13	5,0	4,7	4,7	5,0	5,0	4,7
14	5,0	4,7	5,0	4,0	5,0	5,0
15	5,0	3,7	4,7	5,0	5,0	4,7

APÊNDICE G – Curva Característica dos Itens (CCI)

Curva Característica do item - 11

Curva Característica do item - 12

Curva Característica do item - 13

APÊNDICE H – Dispositivo lançador de aviões de papel

ANEXO A – Modelo de avião Arumakan

1. Fold in half, then unfold.

2. Fold both top corners inward to the center crease.

3. Fold the top point downward.

4. Fold both top corners inward.

5. Fold the bottom point of the middle layers upward to lock the front layers in place.

6. Fold the sloping edges to the center crease to form the wings, then unfold.

7. Remake the keel crease as a mountain fold, then arrange the wings to match the profile.

Arumakan
by David Mitchell

Launch at medium speed.

ANEXO B – Modelo de avião D'Artagnan

Wednesday • January 1, 2014

*New Year's Day
Kwanzaa ends (USA)*

★ Make sure the location of the star on the folding paper corresponds with the star on the directions.

1. Valley fold in half, then unfold.

2. Fold both top corners to the center crease.

3. Mountain fold the top point down.

4. The dotted lines mark the position of the hidden flaps. Fold both top corners to center crease.

5. Fold the sloping edges to the center crease to form the wings, then unfold.

6. Mountain fold the model in half. Arrange the wings to match the flight profile.

Tip: Check your "Folding the Airplanes" page to learn how to mountain and valley fold.

SAVE THIS PAGE.
You'll use it to fold tomorrow's Concorde.

D'Artagnan
by David Mitchell

Launch at medium speed.

ANEXO C – Modelo de avião Glider

1. Fold in half, then unfold.
Fold the top corners inward to the center crease.

2. Fold the top point downward.

3. Fold the tip up.

4. Fold the left-hand corner to the center crease.

5. Mountain fold flap A and tuck A under B.

6. Repeat steps 4 and 5 for the right-hand side. Then mountain fold the plane in half.

7. Fold the wings down.

8. The completed plane should look like this. You may tape the edges under the fuselage and the back of the plane. Bend trailing edge up slightly for best flight.

Glider
by Kyong H. Lee

ANEXO D – Modelo de avião Dart