

UNIVERSIDADE FEDERAL DO PARANÁ

ROBERTO JÚNIOR MARQUES DELFINO VELOSO

PRÁTICAS BASEADAS EM EVIDÊNCIAS E ANÁLISE DO
COMPORTAMENTO: UMA REVISÃO CRÍTICA DA INSTRUÇÃO BASEADA
EM EQUIVALÊNCIA DE ESTÍMULOS NO ENSINO SUPERIOR

CURITIBA

2018

ROBERTO JÚNIOR MARQUES DELFINO VELOSO

PRÁTICAS BASEADAS EM EVIDÊNCIAS E ANÁLISE DO
COMPORTAMENTO: UMA REVISÃO CRÍTICA DA INSTRUÇÃO BASEADA
EM EQUIVALÊNCIA DE ESTÍMULOS NO ENSINO SUPERIOR

Dissertação apresentada como requisito parcial à obtenção de grau de Mestre em Psicologia, no Curso de Pós-Graduação em Psicologia, Setor de Ciências Humanas, da Universidade Federal do Paraná

Orientador: Alexandre Dittrich

CURITIBA

2018

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE
BIBLIOTECAS/UFPR-BIBLIOTECA DE CIÊNCIAS HUMANAS
COM OS DADOS FORNECIDOS PELO AUTOR

Bibliotecária: Rita de Cássia Alves de Souza – CRB9/816

Veloso, Roberto Júnior Marques Delfino

Práticas baseadas em evidências e análise do comportamento: uma revisão crítica da instrução baseada em equivalência de estímulos no ensino superior / Roberto Júnior Marques Delfino Veloso. – Curitiba, 2018.

102 f.

Dissertação (Mestrado) – Universidade Federal do Paraná. Setor de Ciências Humanas, Programa de Pós-Graduação em Psicologia.

Orientador: Prof. Dr. Alexandre Dittrich

1. Comportamento - Avaliação. 2. Psicologia educacional. 3. Comportamento - Modificação. I. Título. II. Universidade Federal do Paraná.

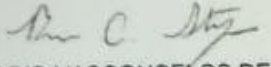
CDD 370.15

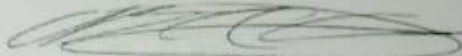
TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em PSICOLOGIA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ROBERTO JUNIOR MARQUES DELFINO VELOSO**, intitulada: **PRÁTICAS BASEADAS EM EVIDÊNCIAS E ANÁLISE DO COMPORTAMENTO: UMA REVISÃO CRÍTICA DA INSTRUÇÃO BASEADA EM EQUIVALÊNCIA DE ESTÍMULOS NO ENSINO SUPERIOR**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua Aprovação no rito de defesa. A outorga do título de Mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 28 de Março de 2018.


BRUNO ANGELO STRAPASSON(UFPR)
(Presidente da Banca Examinadora)

 Participou por
videoconferência
SERGIO VASCONCELOS DE LUNA(PUC/SP)


GABRIEL GOMES DE LUCA(UFPR)

AGRADECIMENTOS

Se cheguei onde estou, isso se deveu à felicidade (para mim) de uma série de acontecimentos fortuitos, como o surgimento deste planeta, da vida que o habitou e da evolução da espécie humana, por exemplo. Diante da magnitude desses acontecimentos, não mantenho zelo pelo mérito de ter feito o que fiz. Cheguei até onde estou, digitando estas palavras, devido a um emaranhado de contingências cuja vastidão se estende ao infinito do tempo.

No entanto, muitas pessoas participaram dessas contingências, e, a despeito do mérito que tiveram pela sorte de acabarmos nos encontrando e participando significativamente um na vida do outro, sinto-me feliz e grato que tantos acasos tenham feito o curso da minha vida trombar com tantas pessoas tão importantes para mim.

A primeira dessas pessoas é meu orientador, prof. dr. Alexandre Dittrich. Mais de cinco anos trabalhando juntos requereram dele paciência incomensurável comigo, e sou muito grato não apenas por isso, mas por toda atenção e dedicação da parte dele em desenvolver em mim comportamentos que me permitiram apresentar a todos vocês esta pesquisa. E não foram só comportamentos acadêmicos, pois consigo perceber a importância dele no meu crescimento e amadurecimento pessoal.

Em segundo lugar, sou infinitamente grato por todos os meus amigos. Uma lista justa de todos eles, suas qualidades e a importância que tiveram para mim, especialmente nestes dois anos de mestrado, levaria muito mais páginas que esta pesquisa. Diante da impossibilidade desse feito, assumirei a injustiça de citar alguns. Meus amigos de infância, Marcelo Avyner, Alex Ricardo e José Geraldo, eu agradeço por terem estado comigo o tempo todo, mesmo que a distância se interpusesse entre nós. A outros amigos, como meu padrinho, conselheiro e parceiro de virar a noite discutindo filosofia, Paulo Roberto; o Bruno Tonet, que nunca recusou estar ao meu lado em nenhuma aventura; a Laura Rabelo, que me adotou e me fez entender o que é ter uma irmã mais velha; e a Fernanda Hirle, a qual tenho a sorte de ter não apenas como amiga e parceira para vida, como também namorada; eu sinto gratidão por me apoiarem de tantas

formas diferentes em momentos importantes e difíceis que me fazem sentir como um monumento de dominós: se um apenas fosse removido, eu haveria desmoronado.

Tenho a sorte de ter uma família extremamente carinhosa e que nunca deixou de me apoiar, incentivar e cuidar de mim por toda essa jornada, ainda que eu tenha desistido de tentar fazê-los compreender o que exatamente eu faço. Queria que todas as pessoas tivessem a sorte de ter uma mãe tão amorosa como a minha, um pai que os apoiasse tanto quanto o meu, um irmão cuja lealdade é inquestionável e uma irmã que encarna a essência da amizade.

Menciono minha gratidão aos professores que compõem a banca que avaliará esta pesquisa. Tive a sorte de ter aprendido com dois deles comportamentos que considero valiosíssimos. Em nome da brevidade, destaco apenas dois desses comportamentos, crucialmente importantes para a execução deste trabalho. Sem a modelagem direta do prof. Bruno Strapasson eu teria tido ainda mais dificuldade para conseguir ler e entender pesquisas experimentais. Quanto ao professor Gabriel de Luca, ele consiste em um bom exemplo de que não tenho mérito algum por nada do que fiz: de todos os elogios que recebi quanto à qualidade da minha redação, sinto que todos eles deveriam ser dirigidos a ele, cuja forma de escrever eu imitei recorrentemente e pela paciência ilimitada para modelar no meu repertório, comportamentos que a gente já deveria ter aprendido na faculdade (e pós-graduação). Tenho comigo que ambos foram figuras importantes e responsáveis pela minha alfabetização científica. Por fim, ao professor Sérgio Vasconcelos de Luna, lamento não ter tido mais oportunidades de convivência e aprendizagem. Sou muito grato pela bondade do professor dispor do seu tempo e paciência para participar desta banca, e, pela sorte das contingências ou pelas consequências de nossas ações, espero corrigir esse déficit de convivência profissional.

“Mas eu não sabia ler. Então meu padrinho teve uma decisão: me enviou para o Currálinho, para ter escola e morar em casa de um amigo dele, Nhô Maroto, cujo Gervásio Lê de Ataíde era o verdadeiro nome social. Bom homem. [...] Vai, acontece, ele me disse: – “Baldo, você carecia mesmo de estudar e tirar carta-de-doutor, porque para cuidar do trivial você jeito não tem. Você não é habilidoso.” Isso que ele me disse me impressionou, que de seguida formei em pergunta, ao Mestre Lucas. Ele me olhou, um tempo – era homem de tão justa regra, e de tão visível correto parecer, que não poupava ninguém: às vezes teve dia de dar em todos os meninos com a palmatória; e mesmo assim nenhum de nós não tinha raiva dele. Assim Mestre Lucas me respondeu: – “É certo. Mas o mais certo de tudo é que um professor de mão-cheia você dava...” E, desde o começo do segundo ano, ele me determinou de ajudar no corrido da instrução, eu explicava aos meninos menores as letras e a tabuada. ”

João Guimarães Rosa, em Grande Sertão: Veredas.

“Por estas e, sem dúvidas, por outras razões, o que se tem ensinado como pedagogia não tem sido uma verdadeira tecnologia do ensino. O ensino nas faculdades, com efeito, não tem sido de modo algum abordado. O professor principiante não recebe preparação profissional. Geralmente começa ensinando simplesmente como foi ensinado e, se melhor, é apenas graças à sua própria e desamparada experiência. O ensino na escola primária e secundária é ministrado principalmente através de ‘estágios’, em que o estudante recebe conselhos e recomendações de professores experimentados. Algumas receitas do ofício e regras práticas são passadas adiante, mas a experiência própria do jovem professor continua a ser a principal fonte de melhora. Mesmo esta modesta tentativa de treinamento de professores tem sido atacada. Tem-se argumentado que o bom professor é simplesmente o que conhece o assunto e está nele interessado. Qualquer conhecimento especial da pedagogia como ciência básica é tido como desnecessário. ”

B. F. Skinner, em Tecnologia do Ensino.

RESUMO

O movimento da educação baseada em evidências propõe que as intervenções realizadas por profissionais da educação sejam escolhidas com base nas melhores evidências disponíveis. A preocupação de analistas do comportamento com a formulação de procedimentos didáticos eficazes pode ser considerada compatível com a proposta da educação baseada em evidências, mas, a despeito disso, suas estratégias de ensino têm sido criticadas por educadores. No âmbito do ensino superior, essas críticas consistem, principalmente, no contraste entre a preferência que analistas do comportamento dão ao ensino de comportamentos mais simples e fáceis de descrever em detrimento das habilidades sofisticadas que se espera de um graduando. Uma estratégia de ensino recente, formulada por analistas do comportamento, apresenta características que podem fazer frente a algumas dessas críticas. Baseada nas tecnologias produzidas pelas pesquisas em equivalência de estímulos, essa estratégia de ensino consiste no uso de procedimentos de formação de classes de equivalência entre estímulos para ensinar comportamentos academicamente relevantes. Bem-sucedida no ensino de repertórios requisito para habilidades mais complexas, como alfabetização e matemática, essa estratégia começou a ser empregada para ensinar repertórios mais complexos, como aqueles exigidos em disciplinas do ensino superior. Considerando esse contexto, o objetivo desta pesquisa consistiu na avaliação das evidências quanto ao uso de procedimentos de formação de classes de equivalência como estratégia didática no ensino superior por meio dos critérios adotados pelas práticas baseadas em evidências na educação. Foi realizada uma busca na literatura e, após a triagem, os artigos encontrados foram avaliados de acordo com os padrões da What Works Clearinghouse, considerada uma importante instituição que publica revisões de práticas baseadas em evidências na educação. Dos 18 artigos que passaram na triagem, apenas três foram aprovados de acordo com os critérios avaliativos, sendo que dois foram aprovados com cautela. O tamanho de efeito calculado a partir do impacto da intervenção foi de $g = 2,18$ e pode ser considerado grande. Os resultados permitiram classificar a instrução baseada em equivalência no ensino superior como apresentando forte evidência de um efeito positivo com nenhuma evidência contrária que a supere, mas com extensão pequena de evidências. Os resultados permitem recomendar, como uma prática baseada em evidências, o uso de instrução baseada em equivalência como estratégia didática no ensino superior. Essa recomendação foi discutida em termos de seus limites, principalmente quanto à quantidade de estudos que compuseram a amostra avaliada. Também foram discutidas as limitações impostas pelos padrões avaliativos da WWC quanto aos resultados desta pesquisa e quanto às práticas baseadas em evidência, como um todo. Por fim, com o sucesso da avaliação de uma estratégia de ensino analítico-comportamental como uma prática baseada em evidência, refletiu-se sobre possíveis intercâmbios entre analistas do comportamento e a comunidade geral de educadores.

Palavras-chave: educação, análise do comportamento, equivalência de estímulos, práticas baseadas em evidências.

ABSTRACT

The evidence-based education movement states that the interventions made by education professionals should be based on the best evidence available. Behavior analysts' concerns with the formulation of effective didactic procedures may be considered compatible with the evidence-based education proposal, but despite this, their teaching strategies have been criticized by educators. In higher education, these criticisms consist mainly of the contrast between the preference that behavior analysts place on teaching behaviors that are simpler and easier to describe, in detriment of the sophisticated skills expected from a graduate student. A recent teaching strategy, formulated by behavioral analysts, has shown features that can address some of these criticisms. Based on the technologies produced by stimulus equivalence research, this teaching strategy consists in the use of procedures to form equivalence classes between stimuli to teach academically relevant behaviors. Successful in teaching repertoires required for more complex skills, such as literacy and math, this strategy began to be employed to teach more complex repertoires such as those required in higher education disciplines. Considering this context, the objective of this research consisted in the evaluation of the evidences regarding the use of procedures of formation of equivalence classes, as a didactic strategy in higher education, through the criteria adopted by evidence-based practices in education. A literature search was performed and, after screening, the articles found were evaluated according to What Works Clearinghouse standards, considered to be an important institution that publishes reviews of evidence-based practices in education. From the 18 articles that passed the screening, only three were approved according to the evaluative criteria, two of which were approved with caution. The effect size calculated from the impact of the intervention was $g = 2.18$, and may be considered large. The results allowed to classify the instruction based on equivalence in higher education as presenting strong evidence of a positive effect with no contrary evidence that overrides it, but with small extent of evidence. The results allow us to recommend, as an evidence-based practice, the use of equivalence-based instruction as a didactic strategy in higher education. This recommendation was discussed in terms of its limits, mainly regarding the number of studies that composed the sample evaluated. The limitations imposed by WWC evaluative standards on the results of this research and on evidence-based practices as a whole were also discussed. Finally, with the successful evaluation of an analytic-behavioral teaching strategy as an evidence-based practice, we reflected on possible exchanges between behavior analysts and the general community of educators.

Keywords: education, behavior analysis, stimulus equivalence, evidence-based practices.

LISTA DE FIGURAS

Figura 3.1 Etapas da avaliação de estudos com delineamento intersujeitos ...	47
Figura 3.2 Critérios para avaliação da elegibilidade de estudos SCD.....	50
Figura 3.3 Exemplo de análise do nível dos pontos de dados em delineamento ABAB.....	51
Figura 3.4 Exemplo de análise da tendência dos pontos de dados em delineamento ABAB	52
Figura 3.5 Exemplo de análise da variabilidade dos pontos de dados em delineamento ABAB	52
Figura 3.6 Exemplo de análise do imediatismo do efeito da intervenção em cada fase	53
Figura 3.7 Exemplo de análise de sobreposição entre as fases	54
Figura 3.8 Exemplo de análise de consistência entre fases similares em um delineamento ABAB	54

LISTA DE TABELAS

Tabela 1.....	56
Tabela 2.....	57
Tabela 3.....	59
Tabela 4.....	73

LISTA DE SIGLAS

PBE – Práticas baseadas em evidências

EBE – Educação baseada em evidência

MBE – Medicina baseada em evidência

AAC – Análise aplicada do comportamento

SCD – Single-Case design (Delineamento de sujeito único)

RCT – Randomized controlled trial (Ensaio randomizado controlado)

QED – Quasi-Experimental design (Delineamento quase-experimental)

WWC – What Works Clearinghouse

EBI – Equivalence based instruction (Instrução baseada em equivalência)

ELB – Equivalência de linha de base

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	Práticas baseadas em evidências na educação.....	14
1.2	A análise do comportamento aplicada e as PBE	17
1.3	A análise do comportamento na educação superior	18
1.4	Instrução Baseada em Equivalência	21
2	CONFLITOS E REFLEXÕES ACERCA DAS PRÁTICAS BASEADAS EM EVIDÊNCIAS E A ANÁLISE DO COMPORTAMENTO	26
2.1	A noção de evidência	28
2.2	Noção de evidência, conhecimento e ciência para o behaviorismo radical.....	30
2.3	Críticas aos delineamentos de grupos.....	36
2.4	O controle divergente exercido pelo movimento das práticas baseadas em evidências	39
3	MÉTODO.....	42
3.1	Identificação da literatura relevante	42
3.2	Triagem	44
3.3	Revisão dos estudos	45
3.3.1	<i>Delineamentos intersujeitos</i>	45
3.3.2	<i>Delineamentos intrasujeitos</i>	47
3.3.3	<i>Síntese dos resultados</i>	56
4	RESULTADOS	62
4.1	Artigos triados.....	62
4.2	Avaliação quanto a elegibilidade a partir dos padrões WWC	72
4.3	Síntese quantitativa dos resultados	75

4.3.1	<i>Tamanho de efeito</i>	76
4.3.2	<i>Índice de aprimoramento</i>	77
4.3.3	<i>Significância estatística</i>	78
4.4	Síntese qualitativa dos resultados	78
5	DISCUSSÃO.....	80
5.1	Recomendação e limitações da EBI como uma prática baseada em evidências	80
5.2	Sobre os limites dos padrões avaliativos adotados pelas PBE	85
5.3	Sobre o uso de delineamentos intersujeitos por analistas do comportamento	87
6	CONCLUSÃO	91
6.1	Limitações e recomendação para futuras pesquisas	93
	REFERÊNCIAS	96

1 1 INTRODUÇÃO

2

3

4 1.1 Práticas baseadas em evidências na educação

5

6

7 O movimento que hoje é conhecido como Práticas Baseadas em
8 Evidências (PBE) foi instituído inicialmente no campo da medicina, na década de
9 1990. Considera-se seu marco inaugural a publicação do *Evidence-Based*
10 *Medicine Working Group* (1992), ao declarar que um novo paradigma neste
11 campo estava emergindo. Vinte anos antes desta publicação, o mesmo grupo já
12 mantinha uma escola de medicina com aprendizagem baseada em problemas,
13 didática considerada inovadora para a época (Eraut, 2007).

14 A prática de basear decisões em evidências, no entanto, é mais antiga.
15 As origens filosóficas das PBE remontam à França, no século XIX (Sacket,
16 1997), em um contexto que pode ser atribuído à tradição filosófica do positivismo.
17 No entanto, há referências a situações com elementos do que hoje se chama de
18 PBE tão antigas quanto os tempos bíblicos (Daniel, 1: 1-16). Tais referências,
19 porém, só surgiram de maneira significativa na literatura científica durante o
20 período do Renascimento, em temas relacionados com testes sanguíneos e o
21 escorbuto (Claridge & Fabian, 2005). Mais recentemente, dois marcos
22 importantes para o desenvolvimento da Medicina Baseada em Evidências (MBE)
23 consistem no trabalho do médico britânico Archibald (Archie) Cochrane e da
24 escola de medicina da McMaster University (Canadá), criada principalmente pela
25 iniciativa do médico canadense David Sackett, em 1967 (Claridge & Fabian,
26 2005). O desenvolvimento paralelo das pesquisas sobre tomada de decisões no
27 campo da administração, na segunda metade do século XX, também influenciou
28 o desenvolvimento da MBE (Eraut, 2007).

29 No contexto da medicina, a noção de PBE é definida de maneiras
30 diferentes, mas complementares. Tradicionalmente, é considerada como “o uso
31 consciente, explícito e sensato da melhor evidência atual na tomada de decisões
32 sobre o cuidado de pacientes” (Sackett, 1997, p. 3), o que significa “...integrar a
33 perícia clínica individual com a apreciação da melhor evidência externa clínica
34 disponível pela pesquisa sistemática” (Sackett, 1997, p. 3). Entretanto, a

35 definição considerada mais completa (Cohen, Stavri, & Hersh, 2004) declara que
36 a MBE é fundamentada em cinco pontos:

37 Primeiramente, decisões clínicas devem ser baseadas na
38 melhor evidência científica disponível; em segundo lugar, o problema
39 clínico, ao invés de hábitos e protocolos, deve determinar o tipo de
40 evidência a ser buscado; em terceiro lugar, identificar a melhor
41 evidência significa usar maneiras de se pensar epidemiológica e
42 estatisticamente; em quarto lugar, conclusões derivadas da
43 identificação e avaliação crítica de evidências são úteis apenas se
44 colocadas em ação no manejo de pacientes ou tomada de decisões de
45 saúde; e, finalmente, o desempenho deve ser constantemente avaliado
46 (Davidoff, Haynes, Sackett, & Smith, 1995, p.1085-1086).
47

48 Um marco fundamental para o movimento da MBE foi a criação da
49 Cochrane Collaboration. A Cochrane Collaboration é uma organização
50 internacional sem fins lucrativos cujo objetivo é ajudar pessoas a tomar decisões
51 bem fundamentadas sobre cuidados de saúde por meio da elaboração e
52 disseminação de revisões sistemáticas de intervenções na área da saúde
53 (Scholten, Clarke, & Hetherington, 2005). Ela foi fundada em 1993, inspirada no
54 trabalho do epidemiologista Archie Cochrane, citado há pouco, cujas pesquisas
55 utilizando testes controlados randomizados (do inglês, *randomized controlled*
56 *trials*, ou RCTs) estabeleceram o paradigma que o movimento da PBE seguiria
57 nos anos seguintes (Claridge & Fabian, 2005).

58 Com o crescimento e difusão da MBE, um movimento semelhante
59 passou a ocorrer na área da educação. A definição da Educação Baseada em
60 Evidências (EBE) é inspirada na da MBE, na medida em que visa integrar a
61 melhor evidência disponível, com o discernimento do profissional e os valores
62 dos clientes para informar a tomada de decisões em todos os níveis de serviços
63 educativos, garantido que estudantes recebam os melhores serviços
64 disponíveis, distinguindo-os de práticas ineficientes (Detrich & Lewis, 2013). Por
65 outro lado, esse papel informativo da EBE é apenas um de seus níveis de
66 operação (Davies, 1999), pois também é necessário que esta venha a
67 estabelecer boas evidências quando as existentes forem escassas ou
68 questionáveis.

69 A aprovação de leis que recomendavam o uso de práticas baseadas em
70 pesquisas científicas contribuiu para impulsionar o movimento da EBE. Políticas
71 como a reforma educacional estadunidense conhecida como *Obey-Porter*
72 *Comprehensive School Reform Demonstration*, de 1997, bem como a lei

73 conhecida como *No Child Left Behind Act*, de 2002, também nos Estados Unidos,
74 tornaram-se famosas por insistir na adoção, por parte de profissionais da
75 educação, de práticas e programas baseados em pesquisas científicas
76 rigorosamente estabelecidas (Slavin, 2002, 2008).

77 Da mesma forma que na área da medicina, instituições de pesquisa
78 foram criadas para apoiar o movimento da EBE. A Campbell Collaboration é uma
79 instituição internacional, de origem britânica, inspirada na Cochrane
80 Collaboration, cujo objetivo é auxiliar profissionais, formuladores de políticas
81 públicas e o público em geral a tomar decisões bem informadas sobre a
82 aplicação de políticas, por meio da elaboração, manutenção e disseminação de
83 revisões sistemáticas de literatura quanto à eficácia de intervenções na
84 educação e em outras áreas, como criminalidade, justiça e bem-estar social
85 (Davies, 2007). A criação da Campbell Collaboration foi inspirada no trabalho de
86 Donald. T. Campbell (1916-1996), um cientista social estadunidense conhecido
87 por defender o uso de métodos científicos para resolver problemas sociais, bem
88 como o uso de testes controlados randomizados como forma de pesquisa
89 científica de boa qualidade (Davies, 2007).

90 Por meio de um empreendimento conjunto entre a Campbell
91 Collaboration e o American Institute for Research, o Departamento de Educação
92 dos Estados Unidos financiou o desenvolvimento da What Works Clearinghouse
93 (WWC), um instituto estadunidense com objetivos idênticos aos da Campbell
94 Collaboration (Davies, 2007). A WWC é considerada o maior dos esforços a fim
95 de sintetizar pesquisas na área da educação, e a mais importante, pois é o único
96 que tem o apoio e financiamento de uma instituição governamental como o
97 Departamento de Educação dos Estados Unidos (Slavin, 2008).

98 No Brasil, o Instituto Alfa e Beto é conhecido como promotor de práticas
99 baseadas em evidências na educação. Criado em 2006, o Instituto é pioneiro no
100 país quanto à promoção e divulgação de pesquisas acerca de temas
101 relacionados à educação, com vistas a subsidiar decisões de profissionais e
102 formuladores de políticas, além do público leigo (Christophe, Elacqua, Martinez
103 e Oliveira, 2015).

104

105 1.2 A análise do comportamento aplicada e as PBE

106

107

108 A preocupação com a fundamentação de práticas em evidências não é
109 recente para analistas do comportamento. Desde cedo em sua obra, Skinner
110 (1953) já apresentava reflexões metodológicas acerca de como estudar
111 cientificamente o comportamento, bem como refletia sobre questões da
112 aplicação do conhecimento adquirido com esse estudo para a solução de
113 problemas humanos. Além disso, considera-se um marco fundamental para a
114 Análise Aplicada do Comportamento (AAC) o artigo seminal de Baer, Wolf e
115 Risley (1968), no qual os autores caracterizaram essa área, conferindo-lhe
116 identidade e prescrevendo algumas práticas que lhe definem.

117 Quase quarenta anos depois, em um manual que se tornou referência
118 para a AAC, Cooper, Heron e Heward (2007), mantiveram suas características
119 definidoras. Ela precisa ser (1) *aplicada*, no sentido de que se dedica a abordar
120 comportamentos de relevância social, de maneira que isso produza melhorias
121 nas condições de vida das pessoas; (2) *comportamental*, no sentido de que seu
122 objeto de pesquisa deve consistir em comportamentos mensuráveis, e não em
123 qualquer correlato possível; (3) *experimental*, na medida em que o pesquisador
124 deve demonstrar a relação funcional entre a variável manipulada e uma mudança
125 significativa em alguma dimensão mensurada do comportamento em questão;
126 (4) *tecnológica*, no sentido de que todos os procedimentos utilizados devem ser
127 descritos com detalhes suficientes para permitir replicação que garanta os
128 mesmos resultados; (5) *conceitualmente sistemática*, no sentido de que os
129 procedimentos produzidos pela pesquisa devem ser descritos em termos dos
130 princípios relevantes a partir dos quais foram derivados; (6) *efetiva*, na medida
131 em que as intervenções projetadas produzam mudanças práticas no
132 comportamento pesquisado, ou seja, as intervenções devem transformar o
133 comportamento-alvo de tal forma que tal mudança seja socialmente relevante;
134 e, por fim, (7) deve proporcionar *generalização*, no sentido de que as mudanças
135 comportamentais produzidas devem persistir após a intervenção, bem como
136 ocorrer em outros cenários que não naqueles nos quais foram inicialmente
137 produzidas, além de se estender para outros comportamentos que não foram
138 diretamente abordados.

139 Todas as dimensões da AAC dizem respeito aos interesses das PBE,
140 com possíveis exceções da dimensão (2) comportamental e (5) conceitualmente
141 sistemática. Pesquisas que não relatam intervenções quanto a problemas
142 aplicados (dimensão 1) sequer entram nas revisões de PBE. O esforço das
143 revisões de literatura consiste, justamente, na tentativa de demonstrar a relação
144 de controle (dimensão 3) entre a intervenção e as mudanças observadas nas
145 pessoas que a receberam, ainda que a forma pela qual isso seja feito tenda a
146 ser distinta nas duas áreas. Não apenas as pesquisas revisadas devem
147 descrever os procedimentos utilizados de maneira clara o suficiente que permita
148 a revisão, como, geralmente, quando o resultado das avaliações é positivo, as
149 práticas revisadas são descritas na forma de manuais e guias práticos, que
150 facilitam sua divulgação (dimensão 4). Os critérios e padrões utilizados para
151 avaliar o resultado das pesquisas são uma forma de se distinguir quais
152 intervenções foram de fato, efetivas (dimensão 6), e, assim o sendo, espera-se
153 que o grupo de pesquisas revisado estabeleça os limites e extensões da
154 aplicação da intervenção (dimensão 7).

155 Se, por um lado, não há um esforço no campo das pesquisas que
156 fundamentam as PBE para que as intervenções sejam estritamente
157 comportamentais (dimensão 3) ou conceitualmente sistemáticas (dimensão 5),
158 como analistas do comportamento prefeririam. Isso se deve à falta de
159 comprometimento das pesquisas que fundamentam as PBE com os princípios
160 filosóficos do Behaviorismo Radical - mas, por outro lado, não há nenhum
161 esforço das PBE em restringir as pesquisas que fundamentam as intervenções
162 que o sejam. Essa compatibilidade com a maioria das características da AAC
163 coloca analistas do comportamento em posição privilegiada para discutir e
164 participar do movimento das PBE. Contudo, como veremos no próximo capítulo,
165 analistas do comportamento ainda não participam integralmente deste
166 movimento.

167

168 **1.3 A análise do comportamento na educação superior**

169

170 A dedicação de analistas do comportamento à área da educação
171 também não é recente. Baseando-se nos princípios comportamentais, analistas
172 do comportamento têm desenvolvido estratégias de ensino que tentam

173 solucionar os problemas decorrentes das formas tradicionais de ensinar e que
174 tornam o ensino mais eficaz. É o caso da estratégia que ficou conhecida como
175 Instrução Programada (PI), caracterizada por demandas claras, em pequenas
176 etapas a serem concluídas pelo aprendiz, cujo comportamento é reforçado de
177 maneira imediata e cujo progresso depende do seu próprio desempenho
178 (Skinner, 1968).

179 Muitas das características da IP tornaram-se princípios propostos para
180 qualquer estratégia de ensino baseada nas leis do comportamento. É preciso
181 garantir que os aprendizes possam aprender no próprio ritmo, e que seus
182 comportamentos sejam consequenciados de maneira imediata; os passos que
183 devem seguir para aprender novos comportamentos devem ser sequenciados
184 do mais fácil ao mais difícil, e ser apresentados em pequenas etapas; é
185 necessário garantir que os aprendizes tenham oportunidades de emitir os
186 comportamentos sendo aprendidos, em vez de confiar que eles emergirão
187 devido apenas à instrução (Kienen, Kubo & Botomé, 2013).

188 Outras estratégias de ensino foram desenvolvidas por analistas do
189 comportamento. Sistemas Personalizados de Ensino (PSI) (Keller, 1968), Ensino
190 de Precisão (*Precision Teaching*) (Lindsley, 1992), e *Interteaching* (Boyce e
191 Hinline, 2002), são alguns exemplos. Ainda que não seja objetivo deste texto
192 recuperar a história dessas propostas ou resumi-las, é preciso apontar que todas
193 essas estratégias de ensino se baseiam nos mesmos princípios
194 comportamentais, ainda que se distingam por um ou outro aspecto que lhes são
195 particulares - mas todas têm, como preocupação fundamental, a eficiência do
196 ensino, e são apoiadas por pesquisas acerca de sua efetividade (Bernstein &
197 Chase, 2013).

198 Essas estratégias de ensino tiveram, cada uma, sua divulgação e
199 destaque, sendo objeto de pesquisa e discussão na comunidade de analistas do
200 comportamento, ainda que não tenham se tornado tendências dominantes na
201 área da educação. Do ponto de vista da educação superior, Bernstein e Chase
202 (2013) afirmam que os principais problemas que estratégias de ensino
203 formuladas por analistas do comportamento têm enfrentado não dizem respeito
204 à efetividade de sua aplicação, mas a outros aspectos.

205 Em relação ao ensino superior, existem três principais fatores pelos
206 quais as contribuições da análise do comportamento têm sido ignoradas por

207 educadores, de acordo com Bernstein e Chase (2013). O primeiro diz respeito a
208 diferenças nas formas de resposta que são avaliadas. Ainda que analistas do
209 comportamento tenham desenvolvido e testado procedimentos para o ensino de
210 repertórios complexos, suas pesquisas sobre educação superior não têm sido
211 bem-sucedidas em deixar claro como as intervenções comportamentais
212 sintetizam, a partir de repertórios mais simples, as habilidades complexas e de
213 alto nível que educadores esperam desenvolver nos estudantes do ensino
214 superior, como solução de problemas complexos, análises críticas, aplicações
215 em contextos novos, sínteses avaliativas, etc.

216 O segundo consiste nas diferenças entre métodos de pesquisa e no tipo
217 de evidência almejada. Pesquisas na área educacional têm preferido o uso de
218 delineamentos de grupo, especialmente aqueles com distribuição aleatorizada
219 de participantes, em detrimento das pesquisas com sujeito único utilizadas
220 majoritariamente na tradição analítico-comportamental. Além disso, como forma
221 de demonstrar controle experimental, analistas do comportamento tendem a
222 preferir investigar respostas mais simples, em ambientes mais controlados, ao
223 invés de repertórios complexos em ambientes mais próximos do natural, como
224 salas de aula do ensino superior.

225 O terceiro fator diz respeito a diferenças entre visão de mundo e
226 pressupostos culturais. A preferência por ensinar comportamentos mais simples,
227 como componentes de desempenhos mais complexos, tende a ir contra a
228 valorização, por parte da maioria dos educadores, de uma perspectiva mais
229 holística - que, quando chega a destacar comportamentos para ser ensinados,
230 ao invés de qualidades abstratas e mentalistas, dá preferência a repertórios
231 complexos, difíceis de se definir com precisão. Analistas do comportamento, ao
232 elaborar estratégias de ensino, acabam por contrariar a tradição que emprega o
233 uso de métodos orais e expositivos, na medida em que utilizam estratégias nas
234 quais o aluno tem um papel mais ativo na aprendizagem.

235 Pode-se acrescentar a tais críticas o fato de que as estratégias de ensino
236 propostas por analistas do comportamento frequentemente requerem bastante
237 esforço dos profissionais do ensino. Tendo em vista o custo requerido para
238 elaborar um programa de ensino de acordo com os princípios propostos por
239 analistas do comportamento, reduzir todo esse trabalho a instrução verbal torna-
240 se muito mais fácil, principalmente quando levamos em consideração outras

241 variáveis que influenciam a escolha e utilização de estratégias de ensino, como
242 a quantidade de aulas que os professores precisam ministrar, os recursos que
243 dispõem, o salário e as gratificações que recebem, por exemplo.

244 As propostas analítico-comportamentais para o ensino não foram
245 adotadas amplamente por educadores e nem mesmo pela própria comunidade
246 proponente. Analistas do comportamento são acusados de não colocar em
247 prática as próprias propostas e de repetir alguns dos erros tradicionais ao
248 ensinar: em casa de ferreiro, espeto de pau (Moreira, 2004). É possível que isso
249 tenha ocorrido devido às mesmas variáveis sob efeito das quais outros
250 profissionais do ensino acabam não adotando estratégias como *interteaching*,
251 PSI e PI, como discutido acima. Contudo, tal explicação não justifica que essa
252 situação continue assim.

253

254 **1.4 Instrução Baseada em Equivalência**

255

256

257 Uma estratégia de ensino recente, contudo, apresenta características
258 que podem fazer frente a algumas das críticas feitas por educadores. Essa
259 estratégia, que tem sido denominada Instrução Baseada em Equivalência (EBI,
260 do inglês *Equivalence Based-Instruction*) (Fields, Travis, Roy, Yadlovker, Aguiar-
261 Rocha e Sturmey, 2009), consiste no uso de contingências de formação de
262 classes de equivalência entre estímulos para o desenvolvimento de
263 comportamentos em contexto de ensino superior. Trata-se de uma estratégia
264 econômica, na medida e que seu uso leva à apresentação, por parte dos
265 aprendizes, de mais comportamentos novos do que foram diretamente
266 ensinados. Trata-se, também, de uma estratégia que desenvolve repertórios
267 complexos, os quais outras áreas chamariam de habilidades cognitivas
268 superiores (De Rose, 1993; Matos, 1999; Rabelo & De Rose, 2015).

269 Proposta inicialmente por Sidman (1994), o estudo da área de formação
270 de classes de equivalências consiste em um abrangente programa de pesquisa
271 a partir do controle de estímulos condicionais. Nos experimentos iniciais, uma
272 resposta era reforçada contingentemente a uma relação de estímulos
273 discriminativos entre as classes de estímulo A e B, por meio de procedimentos
274 de escolha de acordo com o modelo, de maneira que, quando uma segunda

275 contingência entrava em vigor reforçando respostas de escolha entre relações
276 condicionais entre estímulos discriminativos das classes B e C, poderia se
277 observar a emergência do repertório discriminado controlado pela relação entre
278 estímulos das classes A e C, relação esta que não havia sido diretamente
279 treinada.

280 Após suas descobertas iniciais, Sidman & Talby (1982) formularam três
281 critérios para que se verificasse a formação de classes de equivalência. O
282 primeiro critério afirma que as relações devem ser reflexivas - ou seja, diante de
283 uma contingência de pareamento com o modelo, a resposta emitida sob controle
284 de um estímulo A deve também ser emitida sob controle de um estímulo idêntico,
285 A' (a criança cuja resposta vocal “maçã” é emitida diante da figura de uma maçã
286 deve emitir a mesma resposta diante de outra figura idêntica).

287 O segundo critério afirma que as relações devem ser simétricas - ou seja,
288 que a resposta emitida sob controle de uma relação entre estímulos AB, também
289 deve ser emitida sob controle da relação inversa, BA (uma criança que, diante
290 da palavra escrita “maçã”, apontou a figura, deve apontar a palavra “maçã”
291 escrita, quando lhe for apresentada a respectiva figura). Por fim, o terceiro critério
292 refere-se à transitividade das relações - ou seja, quando uma resposta de
293 escolha sob controle da relação entre estímulos AB e uma resposta de escolha
294 sob controle da relação BC são reforçadas, diante de uma relação entre
295 estímulos AC a mesma resposta deve ser emitida (uma criança que, diante de
296 uma maçã, escolhe a palavra “maçã”, e que, diante da palavra “maçã”, fala
297 “maçã”, deve, ao ouvir a palavra maçã, escolher o respectivo objeto dentre
298 outros).¹

299 Um campo de estudos promissor surgiu com as pesquisas sobre
300 equivalência de estímulos, e, juntamente com elas, desenvolveu-se uma
301 tecnologia comportamental aplicada. Inicialmente, essa tecnologia se dedicou ao
302 estabelecimento de repertórios pré-requisitos para habilidades acadêmicas,
303 como a leitura e escrita (e.g. Souza, De Rose, Calcagno, & Galvão, 2004), bem
304 como repertórios matemáticos (e.g. Lorena, Castro-Caneguim, & Carmo, 2013).
305 A partir do sucesso dessas intervenções, recentemente a tecnologia para
306 formação de classes de equivalência foi utilizada para ensinar repertórios em

¹ Sobre o assunto, ver, por exemplo, De Rose (1993).

307 contextos acadêmicos mais complexos, como aqueles exigidos no ensino
308 superior.

309 É possível atribuir às pesquisas conduzidas por Fienup, Critchfield e
310 Covey (2009) e Fields et al. (2009) o início da investigação sobre o uso de EBI
311 como estratégia didática no ensino superior. Em ambas, foi investigado o efeito
312 dessa estratégia na aprendizagem de estatística por alunos do ensino superior,
313 e os resultados reportados apontam a efetividade da intervenção de maneira
314 promissora. Outras pesquisas que se seguiram expandiram os contextos de
315 aplicação para o ensino de outros comportamentos de classes relacionadas a
316 disciplinas de graduação, como neuroanatomia (Fienup, Covey, & Critchfield,
317 2010; Greville, Dymond, & Newton, 2016; Pytte & Fienup, 2012; Reyes-Giordano
318 & Fienup, 2015), pessoas com necessidades especiais (Walker, Rehfeldt, &
319 Ninnes, 2010), delineamentos de sujeito único (Lovett, Rehfeldt, Garcia, &
320 Dunning, 2011; Walker & Rehfeldt, 2012) taxonomia dos operantes verbais
321 clássicos (O'Neill, Rehfeldt, Ninness, Muñoz, & Mellor, 2015) e até mesmo
322 funções do comportamento em contextos aplicados (Albright, Schnell, Reeve &
323 Sidener, 2016). Algumas pesquisas avaliaram a efetividade quando a EBI é
324 aplicada em contextos de sala de aula (Fienup & Critchfield, 2011; Fienup, Mylan,
325 Brodsky, & Pytte, 2016). Outras avaliaram parâmetros da EBI, como o tipo de
326 protocolo de treinamento (Fienup, Wright, & Fields, 2015), ou contribuição de
327 algumas variáveis para sua efetividade (Zinn, Newland, & Ritchie, 2015).

328 Ainda que a maioria das pesquisas apresentadas nos resultados desta
329 revisão indiquem que o uso de EBI é efetivo para ensinar repertórios acadêmicos
330 a estudantes do ensino superior, a qualidade e extensão das evidências
331 apresentadas por essas pesquisas ainda é desconhecida. Diante disso, é
332 possível questionarmos, inspirados pelas PBE, se existem evidências suficientes
333 e de qualidade íntegra que apoiem a adoção dessas práticas pela comunidade
334 de educadores. Como um campo recente, mas promissor, o uso de EBI no
335 ensino superior pode se beneficiar de uma revisão das pesquisas publicadas
336 que, ao agregar e avaliar os resultados já obtidos, informe de maneira concisa à
337 comunidade científica a qualidade e a extensão das evidências produzidas na
338 área.

339 Sendo assim, é objetivo desta pesquisa foi avaliar se o uso de
340 procedimentos de formação de classes de equivalência como estratégia didática

341 no ensino superior pode ser considerado uma prática baseada em evidências de
342 acordo com os critérios da What Works Clearinghouse. Escolhemos a WWC
343 como fonte dos critérios para determinação de PBE devido ao fato de que ela é
344 a maior e mais significativa das instituições de PBE na área da educação (Slavin,
345 2008). Espera-se, com esta pesquisa, fornecer à comunidade científica e
346 profissional uma fonte de informação integrada e sintética acerca das evidências
347 disponíveis quanto a essa estratégia didática.

348 Essa revisão de literatura, no entanto, pode servir a um propósito mais
349 amplo. Considerando o surgimento das PBE na área da medicina, e sua
350 extensão à área da educação, bem como a criação de instituições que
351 promovem revisões sistemáticas de literatura visando estabelecer quais práticas
352 possuem evidências confiáveis e, portanto, deveriam ser utilizadas pelos
353 profissionais do campo da educação em detrimento de outras, constatamos que
354 a preocupação de fundamentar as atividades profissionais a partir de evidências
355 confiáveis é uma tradição promovida na Análise do Comportamento muito
356 anteriormente às discussões recentes.

357 Diante desse contexto, faz-se oportuno aplicar os critérios avaliativos em
358 vigor no campo das PBE a uma estratégia de ensino analítico-comportamental
359 emergente, não apenas a fim de se descobrir como tal estratégia se sairia ao ser
360 avaliada com critérios externos àqueles comuns na análise do comportamento,
361 mas como ocasião para refletir acerca das possibilidades de conjugação e
362 diálogo entre as práticas analítico-comportamentais e as PBE, uma vez que elas
363 têm ganhado maior destaque nos contextos educacionais. Refletir sobre
364 questões como, “por que analistas do comportamento não têm aderido ao
365 movimento das PBE amplamente?” e “quais conflitos entre essas áreas podem
366 estar dificultando que colaborem mais produtivamente?” poderia esclarecer
367 alguns dos problemas envolvidos na situação e, quem sabe, oferecer uma
368 perspectiva de resolução.

369 Antes de iniciarmos a revisão, começaremos pela reflexão proposta
370 acima. Examinaremos de maneira mais aprofundada algumas noções
371 fundamentais para as PBE, assim como a forma que analistas do
372 comportamento lidam com as mesmas noções. Com isso, poderemos comparar
373 as práticas desses dois grupos e tornar mais claros alguns dos conflitos
374 conceituais que os diferenciam. Esperamos, a partir disso, extrair sugestões

375 sobre como uma colaboração mais produtiva pode ser proposta para ambos os
376 grupos.
377

378 2 CONFLITOS E REFLEXÕES ACERCA DAS PRÁTICAS BASEADAS EM
379 EVIDÊNCIAS E A ANÁLISE DO COMPORTAMENTO

380

381

382

383 Para que possam classificar as evidências produzidas pelas pesquisas
384 a partir das quais produzem revisões sistemáticas, as instituições promotoras de
385 PBE adotam critérios avaliativos que geram classificações de tais estudos. Essa
386 classificação dá origem a uma “hierarquia” de evidências (Worral, 2010), uma
387 apreciação acerca de diferentes métodos de pesquisa que seriam considerados
388 melhores que outros. A preocupação consiste em separar as pesquisas com boa
389 validade interna daquelas com problemas e fraquezas metodológicas, e, assim,
390 melhorar a garantia de que os resultados encontrados nas pesquisas decorrem
391 da intervenção promovida pelos pesquisadores, e não de outras variáveis. Nesse
392 sentido, há um consenso, entre essas instituições (Christophe et al., 2015;
393 Higgins, & Green, 2011; Rycroft-Malone, Seers, Titchen, Harvey, Kitson,
394 McCormack, 2004; What Works Clearinghouse, 2017a; Worral, 2010), de que as
395 melhores pesquisas são aquelas que utilizam o delineamento que permite um
396 controle maior quanto a ameaças à validade interna de uma pesquisa.

397 RCTs são delineamentos de grupo, ou inter-sujeitos, nos quais a
398 atribuição de participantes a um ou outro grupo (geralmente grupo experimental
399 e controle), é determinada por algum processo aleatório. Como delineamentos
400 de pesquisa experimental, RCTs tornaram-se o “padrão ouro” de qualidade. O
401 argumento sustentando esse raciocínio é que a aleatorização da distribuição de
402 participantes entre os grupos diminui a formação de tendências em
403 características dos participantes de um grupo ou de outro, que poderiam acabar
404 influenciando os resultados, e, assim, aumentam a fidedignidade da amostra
405 quanto à distribuição daquelas características em relação à população
406 (Shaughnessy, Zechmeister, & Zechmeister, 2012).

407 Delineamentos quase-experimentais (do inglês, *quasi-experimental*
408 *designs* – QED’s), diferenciam-se de RCTs por não atribuírem participantes aos
409 grupos de maneira aleatória. Fora essa característica, não se distinguem:
410 também são pesquisas inter-sujeitos, geralmente usando grupos de pessoas
411 para receber uma intervenção ou não receber nenhuma, nas quais as variáveis

412 de interesse à pesquisa são mensuradas antes e depois da intervenção, em
413 ambos os grupos. No entanto, essa característica diferente dos RCTs coloca
414 QED's em uma posição inferior quanto ao controle de vieses que ameaçam a
415 confiabilidade dos dados (Shaughnessy, Zechmeister, & Zechmeister, 2012).

416 Outros tipos de delineamentos, como aqueles usados em pesquisas
417 qualitativas, sequer entram na hierarquia de evidências, uma vez que não
418 produzem mensurações acerca da eficiência de intervenções e programas. Há
419 discussões acerca da inclusão de resultados de pesquisas qualitativas nos
420 relatórios sobre a efetividade de intervenções na educação (Elliot, 2007;
421 Hammersley, 2007; Hodkinson & Smith, 2007;), mas tal discussão foge dos
422 objetivos deste texto.

423 Delineamentos de sujeito único (do inglês, *single-case design* – SCDs)
424 ocupam um lugar discutível na hierarquia de evidências das instituições que
425 promovem a EBE. Embora se reconheça que esses delineamentos possam
426 produzir boas evidências sobre as relações causais entre variáveis, eles ainda
427 não são tão valorizados como RCTs (What Works Clearinghouse, 2017a; ver o
428 apêndice A). Talvez essa posição seja devida às dificuldades encontradas para
429 transformar os resultados de SCDs na medida padrão adotada para informar a
430 magnitude da aplicação de uma intervenção, conhecida como tamanho de efeito.
431 Apesar dessas dificuldades, pesquisadores têm discutido sobre a adequação de
432 pesquisas que usam SCD em análises estatísticas para meta-análises (Burns,
433 2012; Byiers, Reichle, & Symons, 2012; Horner, Swaminathan, Sugai, &
434 Smolkowski, 2012; Smith 2012).

435 Como visto, o critério utilizado para hierarquizar a qualidade das
436 evidências consiste, principalmente, no controle que o tipo de delineamento
437 oferece em relação a ameaças quanto à validade interna do método, a partir de
438 uma lógica de pesquisa com grupos. Quanto menos dessas ameaças, mais
439 confiáveis são os dados. Contudo, outros critérios também são aplicados, como
440 a ineficácia de pesquisas qualitativas em mensurar o efeito de intervenções. Não
441 parece haver, no entanto, uma avaliação explícita da perspectiva filosófica
442 adotada para o estabelecimento da hierarquia de evidências - pois, como
443 veremos a seguir, aquilo que é considerado uma evidência (ou a melhor
444 evidência) pode variar dependendo do paradigma científico em questão, como é
445 o caso da Análise do Comportamento e do Behaviorismo Radical.

446

447 **2.1 A noção de evidência**

448

449 O que consiste em uma evidência pode variar a depender da perspectiva
450 filosófica adotada. Etimologicamente, a palavra “evidência” está relacionada com
451 experiência, referindo-se àquilo que se apresenta, aos fenômenos que se
452 manifestam, que são óbvios (Rycroft-Malone et al., 2004). Em um sentido mais
453 popular, uma evidência pode ser entendida como uma informação que
454 corrobora, ou sustenta, uma afirmação (Thomas, 2007), ou como alguma
455 garantia conceitual para uma crença ou uma ação (Goldenberg, 2006). Em um
456 sentido filosófico tradicional (derivado da lógica), uma evidência consiste nas
457 informações que portam a verdade ou falsidade de proposições (Thomas, 2007).

458 No paradigma proposto pela MBE, evidência é considerada
459 exclusivamente como evidência de pesquisa, e, implicitamente, evidência
460 produzida por métodos quantitativos de pesquisa (Rycroft-Malone et al. (2004).
461 A noção de evidência adotada pelas PBE é a de que evidências são
462 consideradas fatos que corroboram ou derrubam crenças científicas
463 (Goldenberg, 2006). Evidências também são tratadas como um árbitro neutro e
464 objetivo entre visões, explicações ou teorias competitivas, visando gerar
465 concordância entre observadores (Djulgovic, Guyatt, & Ashcroft, 2009).

466 Essa noção de evidência adotada pelas PBE é compatível com as
467 tradições da filosofia da ciência conhecidas como empirismo e positivismo
468 (Djulgovic, Guyatt, & Ashcroft, 2009; Goldenberg, 2006; Hjørland, 2011). O
469 empirismo é uma tradição filosófica creditada a pensadores britânicos do século
470 XVII, os quais afirmavam ser a experiência o fundamento do conhecimento,
471 vindo esta dos nossos sentidos externos ou da reflexão, entendida como a
472 percepção dos nossos processos mentais (Dutra, 2010).

473 Algumas características da ciência ocidental moderna foram inspiradas
474 pelo empirismo, como o uso de experimentos como forma de adquirir
475 conhecimento confiável sobre o mundo. Nesse aspecto, as PBE defendem um
476 programa empirista na medida em que defendem que os experimentos
477 científicos são a forma privilegiada de conhecimento, e que o uso de estratégias
478 metodológicas mais aperfeiçoadas melhoraria o conhecimento na medida em

479 que melhoraria as experiências que podemos adquirir por meio dos nossos
480 sentidos.

481 O positivismo, por sua vez, diz respeito a duas tradições filosóficas
482 distintas, ainda que semelhantes. O positivismo comtiano, ainda que defenda
483 que o conhecimento necessite de nossa experiência sensorial com os
484 fenômenos que se apresentam para nós (aquilo que é positivo), distingue-se do
485 empirismo ao defender que o conhecimento começa em alguma teoria (e não na
486 experiência), e que esta teoria, ainda que incorreta, orienta as experiências
487 iniciais que a sustentarão ou derrubarão (Dutra, 2005). Segundo essa vertente
488 do positivismo, há estágios diferentes do conhecimento, começando pelo
489 teológico ou fictício, passando pelo metafísico ou abstrato, e chegando ao
490 conhecimento verdadeiro, o positivo ou científico. Esses estágios demarcam
491 formas de conhecer o mundo, e, no estágio mais evoluído do conhecimento,
492 busca-se descrever corretamente as leis da natureza, ou seja, "... as relações
493 invariáveis de sucessão e similaridade dos fenômenos" (Dutra, 2005, p. 52).
494 Nesse sentido, o conhecimento verdadeiro, científico, provém do esforço de
495 tornar as descrições dos fenômenos cada vez mais corretas, e, nesse aspecto,
496 a hierarquia de evidências proposta pelos movimentos de PBE promove a
497 agenda positivista.

498 Outro tipo de positivismo, o lógico, consiste em um sistema filosófico cuja
499 origem é atribuída a um grupo de filósofos e cientistas vienenses, atuantes
500 durante a década de 1920, que receberam a alcunha de Círculo de Viena, e que
501 reconheciam apenas proposições cientificamente verificáveis como significativas
502 (Goldenberg, 2006). Portanto, para positivistas lógicos, qualquer proposição cuja
503 verificação não fosse possível era descartada dos limites daquilo que se
504 considera científico. Para ser verificável, uma proposição deveria ser passível de
505 redução a suas partes elementares, que seriam comparadas com fenômenos no
506 mundo (proposições protocolares). Caso houvesse concordância entre os
507 observadores, então aquela proposição era considerada científica.

508 O esforço de demarcação daquilo que é científico e o que não é,
509 promovido pelo positivismo lógico, é compatível com as propostas da PBE, cujos
510 objetivos, como já vimos, envolvem distinguir tipos de conhecimento diferentes
511 por meio do tipo de evidências que os suportam, apoiando práticas que têm
512 fundamentação em experimentos científicos, de práticas que não possuem tal

513 fundamentação. A noção de evidência das PBE é idêntica à de proposições
514 protocolares, uma vez que também se busca justificar com observações
515 controladas uma proposição, de maneira que se a comunidade científica
516 concordar que tal proposição é compatível com a observação, logo ela é
517 verdadeira.

518 O critério de consenso entre observadores defendido por positivistas
519 lógicos aproxima a noção de evidência adotada por defensores das PBE com
520 esse movimento filosófico. Uma evidência apenas é considerada, na perspectiva
521 das PBE, como válida, se houver concordância por meio da comunidade
522 científica quanto a ela, e, nesse caso, a comunidade consiste principalmente dos
523 pesquisadores que, trabalhando pelas instituições que promovem as PBE
524 (Cochrane Collaboration, Campbell Collaboration, What Works Clearinghouse,
525 etc.), conduzem e divulgam revisões sistemáticas de literatura. Revisões
526 sistemáticas de literatura acabam cumprindo a função de declarar o consenso
527 acerca dos achados de pesquisas científicas e utilizam para isso técnicas de
528 estatística inferencial como chancela para validar o consenso científico.

529 Métodos de pesquisa, como SCD, sobre os quais ainda se discute como
530 integrá-los às revisões sistemáticas de literatura, têm sua credibilidade
531 questionada, independentemente da fidedignidade das evidências que podem
532 produzir. SCDs, no entanto, são considerados de valia entre analistas do
533 comportamento, devido à fidedignidade que podem proporcionar a determinados
534 tipos de pesquisa. A produção de boas evidências pelo uso de SCDs é
535 justificável, por analistas do comportamento, devido a compromissos filosóficos
536 assumidos por tais cientistas, como veremos a seguir.

537

538 **2.2 Noção de evidência, conhecimento e ciência para o behaviorismo** 539 **radical**

540

541

542 A noção de evidência utilizada por analistas do comportamento pode ser
543 encontrada ao se examinar as reflexões filosóficas produzidas e reproduzidas
544 por essa comunidade. O Behaviorismo Radical consiste na filosofia da ciência
545 do comportamento, e, como tal, reúne reflexões acerca do comportamento
546 humano, os meios pelos quais estudá-lo, seus limites e possibilidades de seu

547 estudo, bem como as aplicações e consequências que podem ser derivadas de
548 tal investigação (Skinner, 1974).

549 Parte das reflexões do Behaviorismo Radical podem ser consideradas
550 epistemológicas, pois examinam e discutem sobre a natureza e possibilidades
551 do conhecimento. Para behavioristas radicais, conhecer é comportamento
552 (Moore, 2008; Schnaitter, 1987; Skinner, 1974; Tourinho 1996; Tourinho e Neno,
553 2003; Zuriff, 1980). Há duas interpretações sobre os usos da palavra
554 “conhecimento”, na obra de Skinner, o fundador dessa filosofia. A primeira afirma
555 que conhecimento é comportamento, e que conhecer algo é se comportar de
556 determinada forma diante de estímulos específicos e de maneira a produzir
557 certas consequências (Tourinho, 1996). A segunda interpretação afirma que
558 conhecer é dispor de regras que governam a ação, ou seja, descrições de
559 contingências de reforçamento; neste sentido, o conhecimento não é
560 diretamente o comportamento em contato com a realidade, mas o
561 comportamento sob controle das regras que o tornam eficaz naquela realidade
562 (Tourinho, 1996). Em ambas as interpretações do termo, conhecimento é ação,
563 e não se trata de algo passivo, como uma contemplação ou uma forma de
564 absorver o mundo (Zuriff, 1980).

565 A noção de conhecimento proposta por behavioristas radicais é
566 compatível com a tradição pragmatista da filosofia. Uma grande influência para
567 o pragmatismo do Behaviorismo Radical, tal qual proposto por Skinner, foi a obra
568 de Ernest Mach, que defendida que, ao invés de providenciar uma descrição da
569 realidade, teorias são apenas instrumentos que nos permitem ir de uma
570 observação a outra, sem afirmar que são absolutamente verdadeiras ou
571 absolutamente falsas, mas convenientes ou inconvenientes, efetivas ou
572 ineficientes (Zuriff, 1980).

573 Mach foi uma influência importante para a origem do movimento
574 pragmatista na filosofia estadunidense, mas também exerceu grande influência
575 no surgimento do positivismo lógico. Contudo, há diferenças claras entre
576 positivistas e pragmatistas (Lattal & Laipple, 2003): enquanto estes se
577 preocupam com a concordância entre observadores acerca da verificação das
578 observações, o critério pragmatista de verdade não requer consenso entre
579 observadores, e isso se reflete no Behaviorismo Radical de Skinner (1945); em
580 suas próprias palavras, “o que importa a Robinson Crusóé não é se ele está

581 concordando consigo mesmo mas se ele está conseguindo chegar a algum lugar
582 com seu controle sobre a natureza” (p.552).

583 Essa diferença, marcada pelo critério de verdade proposto pelo
584 pragmatismo, foi bem desenvolvida pelo filósofo estadunidense William James.
585 Para o pragmatismo de James, uma proposição feita por uma pessoa pode ser
586 considerada verdade se, e apenas se, conduzir a uma ação específica e essa
587 ação for bem-sucedida, se levar a uma solução útil (Lattal & Chase, 2003). Para
588 o pragmatismo, proposições verdadeiras possuem valor funcional (Tourinho,
589 1996). Dessa forma, a verdade não é um atributo de uma crença que representa
590 a realidade, mas uma referência à função de uma crença que organiza mais
591 efetivamente a experiência humana, não exclusivamente de maneira prática,
592 mas também intelectual, ou seja, que torna mais efetiva a organização conceitual
593 do mundo (Tourinho & Neno, 2003).

594 É possível identificar semelhanças entre o pragmatismo de James e a
595 epistemologia comportamental de Skinner, ao abordar, ele próprio, o problema
596 da verdade (Skinner, 1974, p.235): “uma proposição é ‘verdadeira’ na medida
597 em que ajuda o ouvinte a responder efetivamente à situação que ela descreve”.
598 Nesse sentido, Skinner (1953) afirma que “respostas a algumas formas de
599 estimulação são mais prováveis de estarem ‘corretas’ que respostas a outras, no
600 sentido de que elas são mais prováveis de conduzir ao comportamento efetivo”
601 (p.139). Esse critério de verdade, tal como formulado, possibilita a interpretação
602 de níveis de verdade graduais, em contraposição a verdades absolutas, e, sendo
603 assim, aproxima-se da proposta de hierarquia de evidências das PBE, mas não
604 pelos mesmos motivos: para as PBE, as melhores evidências vêm de certas
605 estratégias metodológicas que permitem uma descrição de eventos mais fiel à
606 realidade (perspectiva positivista), enquanto que, como já foi abordado, para o
607 critério pragmático, a propriedade de verdade de uma proposição varia conforme
608 sua efetividade prática, as consequências a que leva, e não ao método
609 empregado, ou seja, a forma de sua descoberta.

610 A noção pragmatista do conhecimento estende-se também ao
611 conhecimento científico. Para o pragmatismo de Peirce, James e Dewey, a
612 atividade dos cientistas é produzir descrições da natureza que proporcionem
613 interações mais efetivas com a natureza (Tourinho, 1996). Explicações
614 verdadeiras são aquelas que auxiliam o cientista a lidar com o comportamento

615 de maneira produtiva, ou que ajudam alguém a resolver um problema de maneira
616 eficiente (Tourinho & Neno, 2003).

617 Na perspectiva do Behaviorismo Radical, Ciência é aquilo que os
618 cientistas fazem. Não se trata de algo independente dos cientistas, mas de seu
619 próprio comportamento enquanto trabalham, e, por causa disso, para
620 entendermos como a ciência funciona, devemos examinar aquilo que os
621 cientistas fazem ao agir como cientistas (Creel, 1987). De acordo com Skinner
622 (1953), “o ‘sistema científico’ é projetado para nos capacitar a lidar com um
623 assunto mais eficientemente (p. 14)”. Esse sistema seria organizado na forma
624 de “... regras para a ação efetiva, e há um sentido especial pelo qual o
625 conhecimento seria “verdadeiro” se ele compreende a ação mais efetiva possível
626 (Skinner, 1974, p. 235)”. Como regras, as leis formuladas pelos cientistas não
627 são obedecidas pela natureza, pois, como produtos verbais, elas têm efeito nas
628 pessoas devidamente treinadas para segui-las, e, assim, atingir um controle
629 melhor sobre o mundo ao seu redor (Skinner, 1974).

630 Há de se reconhecer que há um esforço da comunidade verbal em
631 colocar o comportamento do cientista sob controle estrito dos estímulos que
632 compõem o fenômeno que ele se dedica a estudar (Tourinho, 1996). Parte desse
633 esforço reflete-se nas práticas adotadas pela comunidade científica, como as
634 várias metodologias de pesquisa ou mesmo os rituais e trâmites para publicação
635 de resultados de investigações científicas. Isso não resulta, no entanto, em dizer
636 que o comportamento do cientista é mais fiel à realidade do que o
637 comportamento do leigo: o comportamento de ambos é modelado pelo ambiente,
638 ainda que cada um a seu modo (Skinner, 1974).

639 A partir de uma perspectiva pragmatista do comportamento dos
640 cientistas, pode-se propor objetivos para a ciência. De acordo com Skinner
641 (1953), a ciência, como tentativa de descobrir relações sistemáticas entre
642 eventos, preocupa-se em prever e controlar os fenômenos que estuda. Predição
643 e controle consistem em uma especificação da perspectiva pragmática do
644 comportamento científico, diferenciando explicações mais efetivas do
645 comportamento, ou seja, aquelas que auxiliam os cientistas a agir de maneira
646 mais efetiva, a resolver problemas (Tourinho & Neno, 2003).

647 Os cientistas atingem seu objetivo de prever e controlar seu objeto de
648 estudo ao descrever relações funcionais entre os fenômenos estudados.

649 Relações funcionais consistem em proposições acerca de como certos eventos
650 tendem a ocorrer em uma certa ordem (Skinner, 1953). Em relação à
651 investigação científica do comportamento humano, Skinner (1953) sugere que
652 os objetivos da ciência são atingidos ao se descrever as relações funcionais
653 entre o comportamento dos organismos e o ambiente:

654 As variáveis externas das quais o comportamento é função
655 provêm aquilo que pode ser chamado de análise casual ou funcional.
656 Nos comprometemos a prever e controlar o comportamento do
657 organismo individual. Essa é nossa “variável dependente” – o efeito
658 pelo qual procuramos a causa. Nossas “variáveis independentes” - as
659 causas do comportamento – são condições externas das quais o
660 comportamento é uma função. Relações entre as duas – as “relações
661 de causa e efeito” no comportamento – são as leis da ciência. Uma
662 síntese dessas leis expressas em termos quantitativos produz uma
663 imagem abrangente do organismo como um sistema se comportando.
664 (Skinner, 1953, p. 35)

665

666

667 Durante o desenvolvimento da Análise do Comportamento,
668 delineamentos do tipo sujeito-único emergiram como estratégia de pesquisa por
669 meio das quais é possível obter descrições de relações funcionais entre eventos
670 do ambiente e o comportamento dos organismos, e, assim, atingir a previsão e
671 controle almejados pelo empreendimento científico. SCDs consistem em
672 estratégias de pesquisa nas quais o comportamento de um indivíduo é medido
673 repetidas vezes, utilizando o mesmo instrumento de medida, sob duas ou mais
674 condições diferentes (Iversen, 2013), de maneira que seja possível constatar
675 regularidades no comportamento nas diversas fases e comparar as mudanças
676 com a administração e retirada de uma intervenção.

677 Delineamentos de sujeito-único possuem diversas formas (Iversen,
678 2013). A mais simples consiste em uma fase de medidas de linha de base e outra
679 de medidas de intervenção, quando a variável independente é manipulada,
680 conhecido como delineamento A-B. Uma forma mais completa do delineamento
681 A-B envolve adicionar mais uma fase, na qual a manipulação da variável
682 independente é revertida - isto é, a intervenção é removida (delineamento A-B-
683 A). Adicionar mais fases nas quais a variável independente é apresentada e
684 removida cria delineamentos A-B-A-B...n, nos quais a repetição da apresentação
685 e remoção da variável independente demonstra controle experimental.

686 Quando é impossível, extenuante ou antiético reverter os efeitos da
687 apresentação da variável independente sobre o comportamento dos sujeitos ou

688 participantes da pesquisa, um delineamento de linha de base múltipla pode ser
689 preferível (Iversen, 2013). Nesse tipo de delineamento, a fase de reversão não
690 ocorre, mas, alternativamente, compara-se as medidas obtidas de linhas de base
691 que estão ocorrendo simultaneamente (diferentes classes de respostas), mas
692 nas quais a variável independente é manipulada em momentos diferentes.

693 Outras formas de delineamentos de sujeito-único são possíveis, como,
694 por exemplo, delineamentos de mudança de critérios, alternância de tratamentos
695 e multi-elementos. No entanto, uma das características cruciais de todos os
696 delineamentos de sujeito-único consiste na replicação. Tipos de replicação
697 podem ser baseados no próprio sujeito, ou usando outros sujeitos. Quando se
698 compara o comportamento de um participante em fases diferentes, dois tipos de
699 replicação são possíveis: “(a) a estabilidade do comportamento do indivíduo de
700 uma observação para a próxima sob circunstâncias constantes dentre de uma
701 condição e (b) a estabilidade de uma mudança no comportamento do indivíduo
702 de uma condição experimental à outra” (Perone & Hursh, 2013, p.123). Ao
703 acrescentarmos mais fases nas quais se manipula, ou se reverte a manipulação
704 da variável independente enquanto continuamos a medir o comportamento,
705 estamos, portanto, criando replicações. A replicação também pode ocorrer por
706 meio da utilização de outros participantes na pesquisa. Ela pode ser direta ou
707 sistemática: direta quando se repetem as condições do experimento com outros
708 participantes, e sistemática quando se repete o experimento, mas com variações
709 em suas condições (Perone & Hursh, 2013).

710 Replicação é uma importante forma de demonstrar confiabilidade dos
711 resultados experimentais (Branch & Pennypacker, 2013; Perone & Hursh, 2013).
712 Por meio da repetição da manipulação da variável independente, apresentando
713 e removendo-a, enquanto se continua a mensuração da variável dependente, é
714 possível afirmar de maneira mais segura que as mudanças na variável
715 dependente são decorrentes daquelas na variável independente. Se, por meio
716 de replicação sistemática, esse mesmo raciocínio for aplicado em condições nas
717 quais certas variáveis mudam, e caso se observe a mesma mudança na variável
718 dependente em função da variável independente, então a relação entre as duas
719 variáveis ganha ainda mais confiabilidade. Quanto mais replicações repetem o
720 mesmo efeito de uma variável em função da outra, melhor se demonstra a
721 relação de dependência entre elas, e não em função de outras variáveis. Dessa

722 forma, por meio da replicação podem ser obtidas evidências confiáveis do efeito
723 de uma intervenção no comportamento dos indivíduos.

724 Analistas do comportamento têm confiado na eficácia de delineamentos
725 de sujeito-único como a melhor forma de se obter evidências acerca do
726 comportamento dos organismos. Isso implica que, segundo a perspectiva do
727 Behaviorismo Radical, delineamentos de sujeito-único têm permitido analistas
728 do comportamento descrever relações de dependência entre certas variáveis e
729 o comportamento sendo estudado de maneira que lhes possibilita prever e
730 controlar seu objeto de estudo, formulando leis científicas que tornam a ação
731 daqueles que a utilizam mais efetiva para lidar com o comportamento de pessoas
732 ou outros animais, ou ainda resolver problemas relacionados a ele.

733 Outro motivo pelo qual analistas do comportamento têm preferido
734 delineamentos de sujeito-único diz respeito a características de outros tipos de
735 delineamento. De acordo com a perspectiva do Behaviorismo Radical, outros
736 tipos de delineamento populares na Psicologia, como os delineamentos de
737 grupo, possuem características que interferem na previsão e controle
738 proporcionados pelos resultados de tais experimentos de maneira que tornam as
739 relações descritas por eles não tão efetivas quanto aquelas proporcionadas
740 pelos resultados de delineamentos de sujeito-único.

741

742 **2.3 Críticas aos delineamentos de grupos**

743

744 O conhecimento científico produzido pelo uso de estratégias de pesquisa
745 com grupos tem sua eficiência, ao controlar o comportamento daqueles que o
746 empregam, prejudicada por, no mínimo, dois fatores. O primeiro fator diz respeito
747 à representatividade da amostra da população. Delineamentos que utilizam
748 grupos de pessoas o fazem como forma de inferir que é provável que uma
749 determinada amostra retirada da população sendo estudada possui as mesmas
750 características, em média, que todo o restante da população (Shaughnessy,
751 Zechmeister, & Zechmeister, 2012).

752 Ao se mensurar a média do valor de determinada variável nos
753 participantes da amostra, é possível confiar que aquele valor representa a média
754 de toda a população de interesse. No entanto, o valor médio de uma determinada
755 variável obtida de uma amostra pode resultar de distribuições substancialmente

756 diferentes dos escores do grupo, e, dessa forma, não proporcionam um indicativo
757 confiável do valor daquela variável para os indivíduos do grupo (Branch &
758 Pennypacker, 2013). Se todos os participantes de um grupo apresentam
759 desempenho alto, ou baixo, a média do desempenho será alta, ou baixa,
760 respectivamente. Contudo, basta um resultado drasticamente alto para tornar
761 médio o desempenho de um grupo de indivíduos com resultados baixos. É
762 possível que valores nem altos, nem baixos, sejam obtidos por meio de diversas
763 combinações do desempenho dos indivíduos, e, no final, o resultado dessas
764 combinações, agregado na média do desempenho grupo, pode não se
765 representativo de nenhum dos participantes da amostra, e, assim confundir ou
766 enganar os resultados.

767 Portanto, ao se mensurar o efeito de uma variável independente sobre o
768 comportamento das pessoas em um grupo, é possível que a relação
769 demonstrada não se repita da mesma maneira, ou na mesma direção, para as
770 pessoas dentro daquela população. O problema, em uma perspectiva
771 behaviorista radical, não consiste na presença ou ausência do efeito, mas no
772 risco que a probabilidade daquele efeito não ser aplicado a alguém que não foi
773 representado pela amostra oferece ao critério pragmático de verdade. Um
774 profissional que está agindo sob controle de uma descoberta científica obtida por
775 amostragem pode não ser tão efetivo quanto o necessário, ou possível, ao lidar
776 com aquela população, pois existe uma probabilidade de que o indivíduo com o
777 qual está interagindo não seja representado pelo desempenho da amostra
778 utilizada na pesquisa.

779 O segundo fator que pode prejudicar os delineamentos de pesquisa com
780 grupos consiste, novamente, no uso de estratégias de estatística inferencial para
781 lidar com os resultados da manipulação da variável independente, sendo o
782 principal desses problemas a lógica utilizada pelo teste de hipóteses. De acordo
783 com a lógica dessa estratégia (Branch & Pennypacker, 2013), ao se avaliar a
784 relação funcional entre uma variável independente (uma intervenção), sobre uma
785 variável dependente (o desempenho de pessoas), propõe-se duas hipóteses
786 antes da coleta dos dados: (a) não há diferença no desempenho entre o grupo
787 que recebeu a intervenção e o grupo que não recebeu – a hipótese nula, e, (b)
788 há diferença no desempenho entre o grupo que recebeu a intervenção e o grupo
789 que não recebeu – a hipótese experimental ou de trabalho.

790 Com a coleta dos dados após a administração da intervenção, conduz-
791 se testes de significância estatística para determinar qual a probabilidade de, se
792 a hipótese nula for verdadeira, os dados terem sido obtidos devido a outros
793 fatores que não foram estabelecidos na pesquisa - ou seja, a fatores estranhos
794 (Branch & Pennypacker, 2013). Caso o resultado do teste de significância
795 estatística, utilizando os dados observados, mostre que há pouca probabilidade
796 ($p < 0,05$) de que aqueles resultados foram obtidos ao acaso, então assume-se
797 que a hipótese nula está incorreta e que a hipótese experimental é verdadeira.
798 O erro consiste na suposição de que, ao se mostrar que é pouco provável que a
799 hipótese nula seja verdadeira, logo a hipótese experimental está correta e a
800 diferença no desempenho entre grupo experimental e grupo controle não
801 decorreu do acaso, mas da intervenção (Branch & Pennypacker, 2013).

802 Ao se descobrir que é pouco provável que a diferença observada nos
803 dados não foi obtida ao acaso, e, portanto, que a hipótese nula não deve ser
804 adotada, mostra-se apenas que a hipótese nula não deve ser adotada: não se
805 segue logicamente que a hipótese alternativa está correta. Os testes de
806 significância estatística não apresentam informações acerca da veracidade da
807 hipótese experimental (Branch & Pennypacker, 2013). Quando assumimos a
808 verdade da hipótese experimental por meio desse artifício estatístico estamos
809 inferindo um resultado a partir dos dados. Essa inferência baseia-se em uma
810 medida probabilística e, devido a isso, prejudica a confiabilidade da informação
811 obtida. Em outras palavras, apostamos que há uma baixa probabilidade (meio
812 por cento) de que aqueles dados foram obtidos ao acaso, portanto, nossa
813 hipótese experimental está correta. A partir de uma perspectiva behaviorista
814 radical, agir sob controle de uma informação inferida, de certeza provável, pode
815 prejudicar a efetividade com que alguém, comportando-se sob controle dessa
816 descoberta científica, aplica alguma intervenção nela baseada - e, assim,
817 ameaça o critério pragmático da verdade.

818 Diante dos prejuízos que o uso de amostragem estatística e teste de
819 hipóteses oferecem em comparação ao uso de delineamentos de sujeito único
820 para a ação daqueles que se comportam sob controle do conhecimento
821 produzido pelo uso dessas estratégias, torna-se claro porque analistas do
822 comportamento podem entrar em conflito com as práticas adotadas pelos
823 proponentes das PBE - que, como vimos, elegem delineamentos de grupo, ou

824 melhor, RCTs, como a melhor evidência disponível. Para analistas do
825 comportamento, os resultados de RCTs podem ser informativos, mas o
826 comportamento controlado por esses achados pode ter sua efetividade
827 prejudicada, em comparação com o comportamento que fica sob controle de
828 achados científicos por meio do uso de SCDs. As melhores evidências, de
829 acordo com a lógica do Behaviorismo Radical, vêm das várias formas de
830 replicação. Esse conflito entre práticas de grupos culturais, aqui exposto, pode,
831 no entanto, decorrer de outra questão: o que realmente significam práticas
832 baseadas em evidências?

833

834 **2.4 O controle divergente exercido pelo movimento das práticas** 835 **baseadas em evidências**

836

837 Quando se questiona se uma prática é baseada em evidências, duas
838 questões diferentes se confundem. A primeira diz respeito à fundamentação de
839 uma descoberta científica, e pode ser exemplificada assim: existem evidências
840 sólidas de que uma intervenção x é responsável pela mudança y na população?
841 Como alguns autores apontam (Goldenberg, 2006; Hammersley, 2007; Rycroft-
842 Malone et al., 2004), trata-se de uma indagação com validade óbvia, uma vez
843 que claramente devemos adotar práticas que se mostraram efetivas mesmo sob
844 controle experimental rígido. Que profissional se oporia a usar as melhores
845 evidências disponíveis ao lidar com seus clientes? Ao se questionar acerca da
846 fundamentação de uma descoberta científica, exerce-se o ceticismo quanto à
847 estabilidade da relação de dependência entre duas variáveis - e, como vimos
848 anteriormente, delineamentos de sujeito-único podem responder a esse tipo de
849 pergunta sem os vieses probabilísticos de delineamentos de grupo. Mas, ainda
850 assim, delineamentos de grupo têm sido preferidos.

851 O argumento aqui defendido é que esse questionamento tem sido
852 confundido com outro. Ao se questionar se uma prática é baseada em
853 evidências, tem sido defendido que aquelas práticas cujas pesquisas científicas
854 mostram que possuem maiores efeitos na população devem ser preferivelmente
855 adotadas pelos profissionais, em relação a práticas cujas pesquisas mostram
856 que os efeitos na população são menores, inexistentes, ou para as quais não
857 exista informação confiável disponível. Isso não é a mesma coisa que questionar

858 se existem evidências confiáveis acerca das relações causais (funcionais) entre
859 a intervenção e a mudança produzida por ela.

860 Essa questão acerca das evidências quanto à extensão do efeito de uma
861 intervenção sobre uma população remete não à fundamentação de uma
862 descoberta científica, mas ao âmbito das ciências atuariais. Ciências atuarias
863 consistem no campo de estudo e desenvolvimento de técnicas quanto aos riscos
864 e expectativas relacionados a grupos, sendo bastante empregadas para o
865 desenvolvimento de seguros ou para formulação de políticas públicas. Portanto,
866 ao se questionar a evidência de uma intervenção, de maneira que seja possível
867 descobrir que impacto é possível esperar do uso daquela intervenção na
868 população, em detrimento de outras intervenções, não estamos questionando
869 sua fundamentação em um sentido exclusivamente epistemológico, ou seja, não
870 questionamos se estamos seguros de que há uma relação de dependência entre
871 a aplicação da intervenção e a mudança em determinada característica da
872 população, mas quanta melhora essa intervenção pode proporcionar aos grupos
873 que a receberão, qual a proporção entre o tamanho do efeito da intervenção e
874 os custos de sua aplicação em relação ao outras intervenções, quão
875 abrangentes são seus resultados em relação à população – qual a probabilidade
876 de que mais pessoas da população sejam beneficiadas por aquela intervenção,
877 ou seja, a representatividade da amostra.

878 No entanto, quando analistas do comportamento são questionados se
879 suas práticas são baseadas em evidências, é mais provável que respondam ao
880 primeiro sentido da questão (epistemológico) do que o segundo (atuarial). O
881 controle divergente (quando respostas diferentes e de mesma probabilidade
882 estão sob controle de um mesmo estímulo) exercido por essa pergunta, da forma
883 como é feita, pode ser a causa de boa parte da polêmica que os proponentes
884 das PBE têm causado no âmbito científico (e.g. Thomas & Pring, 2007, acerca
885 da reação de educadores tentando responder às duas questões de maneira
886 indistinta).

887 Ao se questionar, portanto, se uma intervenção realizada por analistas
888 do comportamento possui evidências, não se trata exclusivamente de
889 demonstrar relações funcionais entre variáveis. Isso configura uma questão para
890 a Análise do Comportamento responder: sob uma perspectiva atuarial, quais
891 evidências dispomos acerca da efetividade das intervenções comportamentais?

892 Analistas do comportamento têm criticado o uso de delineamentos de grupo para
893 constatar relações funcionais por vários motivos (alguns dos quais apresentados
894 anteriormente), mas alguns concedem que essas estratégias têm valor
895 reconhecido para responder a importantes questões atuariais (Branch &
896 Pennypacker, 2013; Iversen, 2013; Smith, 2013).

897 Smith (2013) narra um caso exemplar para ilustrar esse ponto: o
898 desenvolvimento de procedimentos para ensinar mulheres a realizar o
899 autoexame de mama - uma intervenção criada por analistas do comportamento,
900 que passou por refinamentos e demonstrou ser efetiva sob condições
901 controladas em laboratório e, posteriormente, sob condições próximas àquelas
902 encontradas fora do laboratório. Em uma perspectiva analítico-comportamental,
903 as evidências eram claras acerca da efetividade da intervenção. No entanto:

904 ...o status do autoexame de mama como baseado em
905 evidências depende de uma questão atuarial quintessencial: ele [o
906 autoexame] previne mortes devido ao câncer de mama? A revisão da
907 [Colaboração] Cochrane examinou os RCTs que abordavam essa
908 questão e descobriu-se que a resposta era não, talvez porque aprender
909 um procedimento discreto de autoexame falha em abordar as barreiras
910 de se realizar os exames consistentemente ou com precisão. (Smith,
911 2013, pp. 21-22)

912

913 Diante da pergunta acerca das evidências que baseiam uma prática,
914 analistas do comportamento estão confortáveis quanto a capacidade que seus
915 métodos preferidos de pesquisa possuem para demonstrar relações funcionais.
916 As diversas replicações realizadas demonstram boas evidências dos processos
917 comportamentais básicos e algumas de suas aplicações para resolver
918 problemas práticos da sociedade. Contudo, com a popularização das PBE em
919 muitos dos campos de interesse aplicado da Análise do Comportamento, e com
920 o esclarecimento acerca da natureza atuarial da questão das evidências que
921 fundamentam suas práticas, é chegada a hora de analistas do comportamento
922 apresentarem as evidências acerca da efetividade das tecnologias
923 comportamentais sobre comportamento da população. Enquanto apreciações
924 nesse sentido têm sido feitas recentemente no campo da psicoterapia (Leonardi,
925 2016), cabe aos interesses desta pesquisa estender esse tipo de análise ao
926 campo da educação.

927

928 3 MÉTODO

929

930

931 Foi escolhido o modelo para revisões sistemáticas proposto pela *What*
932 *Works Clearinghouse* (2017a; 2017b) como método de desenvolvimento desta
933 pesquisa, o qual será apresentado neste capítulo. O modelo dessa instituição foi
934 escolhido devido ao fato de ela ser a maior e mais significativa (Slavin, 2008) das
935 iniciativas de avaliação de PBE na área da educação.

936 O modelo da WWC é detalhado em dois manuais disponibilizados
937 gratuitamente no site da instituição. Esse modelo é descrito em duas obras, o
938 Manual de Procedimentos da WWC (WWC, 2017a), que apresenta as cinco
939 etapas da revisão sistemática propostos pela instituição, e o Manual de Padrões
940 da WWC (WWC, 2017b), que descreve os padrões pelos quais as pesquisas
941 revisadas devem ser avaliadas.

942 De acordo com a WWC (2017a), uma revisão sistemática de literatura
943 deve transcorrer em cinco etapas: (1) desenvolvimento do protocolo de revisão,
944 o qual especifica os parâmetros de busca dos estudos e quais os critérios para
945 incluir estudos nos resultados; (2) identificação da literatura relevante, que
946 consiste na busca de estudos nos bancos de dados, e outras fontes, como
947 estabelecido no protocolo; (3) triagem dos estudos, etapa na qual os estudos
948 identificados na etapa anterior são revisados para se determinar se são elegíveis
949 dentro dos parâmetros estabelecidos no protocolo da pesquisa; (4) revisão dos
950 estudos, que consiste na avaliação dos estudos triados quanto aos padrões da
951 WWC; e (5) relatório dos resultados, que se trata da integração dos resultados
952 obtidos pela revisão em sumários de fácil compreensão.

953 Nos tópicos a seguir, apresentamos o protocolo de revisão adotado
954 nesta pesquisa, apontando os critérios de identificação da literatura relevante,
955 bem como os critérios de triagem, revisão e relato dos resultados.

956

957

958 **3.1 Identificação da literatura relevante**

959

960

961 Para a busca dos textos, foram utilizados os seguintes bancos de dados,
962 sugeridos pela WWC (2017a) para o desenvolvimento de pesquisas dentro de
963 seus padrões:

- 964 • Academic Search Premier
- 965 • Campbell Collaboration
- 966 • Dissertation Abstracts
- 967 • EconLit
- 968 • Education Reserach Complete
- 969 • EJS E-Journals
- 970 • ERIC
- 971 • SocINDEX with Full Text
- 972 • ProQuest Dissertations & Theses
- 973 • PsycINFO
- 974 • SAGE Journals Online
- 975 • Scopus
- 976 • WorldCat

977 Além desses, os seguintes foram incluídos:

- 978 • SciELO: um banco de dados desenvolvido em colaboração por
979 universidades latino-americanas, que agrega publicações
980 científicas latino-americanas e caribenhas.
- 981 • BVS-Psi: desenvolvida pelo Conselho Federal de Psicologia, em
982 parceria com a Pontifícia Universidade Católica de Campinas, a
983 Biblioteca do Instituto de Psicologia da Universidade de São
984 Paulo e a Biblioteca Virtual de Saúde, a qual reúne diversos
985 bancos de dados nacionais e internacionais, focados em
986 universidades latino-americanas.

987 As palavras-chave utilizadas na busca, em bancos de dados
988 internacionais, foram “equivalence-based instruction”, “equivalence-based
989 teaching” e “equivalence-based learning”. Nos bancos de dados nacionais, foram
990 utilizadas as palavras-chave “ensino baseado em equivalência”, “instrução
991 baseada em equivalência”, e “aprendizagem baseada em equivalência”. Além
992 disso, devido a distinções de nomenclatura entre aquela praticada
993 internacionalmente e aquela praticada no Brasil, nos bancos de dados nacionais,

994 também foram realizadas buscas com os seguintes termos: “equivalência de
995 estímulos” e “ensino superior”; “equivalência de estímulos” e “alunos de
996 graduação”; “equivalência de estímulos” e “faculdade”. Caso esses termos não
997 produzissem resultados nos bancos de dados nacionais, uma terceira alternativa
998 era utilizada, com termos mais amplos: “equivalência de estímulos” e “instrução”;
999 “equivalência de estímulos” e “ensino”; “equivalência de estímulos” e
1000 “aprendizagem”.

1001 Por fim, com a leitura dos artigos revisados, foram buscados artigos nas
1002 referências dos estudos que pudessem participar da amostra, de acordo com os
1003 critérios de triagem.

1004

1005

1006 **3.2 Triagem**

1007

1008

1009 Foram conduzidos dois tipos de triagem dos textos. Uma triagem inicial
1010 consistiu na exclusão dos textos, que, por não apresentarem os critérios de
1011 inclusão (descritos abaixo), não estão no escopo do interesse desta revisão. Em
1012 seguida, foi realizada uma triagem de elegibilidade, de acordo com os padrões
1013 da WWC (2017a).

1014 Os critérios de inclusão a seguir foram aplicados nos textos buscados,
1015 por meio da leitura do título, resumo e dados gerais de publicação. Textos que
1016 não incluíam todos estes critérios não fizeram parte desta revisão:

1017 - A data de publicação compreendeu o intervalo dos últimos vinte anos
1018 (1997 a 2017). Esse intervalo de tempo é estipulado pela WWC (2017a) e pode
1019 consistir no período em que certos resultados de pesquisas científicas podem se
1020 tornar obsoletos frente a novas descobertas.

1021 - Pesquisas que relataram a mensuração dos efeitos de uma intervenção
1022 no desenvolvimento de repertórios comportamentais do público-alvo.

1023 - O público-alvo de alunos do ensino superior, não obstante a idade.

1024 - Intervenção que consistiu em algum procedimento de formação de
1025 classes de equivalência comportamental.

1026 - Os repertórios comportamentais desenvolvidos deverão consistir nos
1027 comportamentos-objetivo de ensino tipicamente ensinado em contexto da
1028 educação superior.

1029 A segunda triagem avaliou os artigos quanto à sua elegibilidade, de
1030 acordo com os padrões da WWC (2017b). Foram considerados apenas estudos
1031 que utilizassem delineamentos de pesquisa experimental, intersujeitos (RCTs e
1032 QED's) ou intrasujeitos (SCDs). Dessa forma, estudos com delineamentos de
1033 levantamento ou caracterização, como revisões de literatura, avaliações
1034 qualitativas dos impactos das intervenções ou meta-análises, não participaram
1035 dos resultados.

1036

1037

1038 **3.3 Revisão dos estudos**

1039

1040

1041 Os estudos resultantes da triagem foram lidos novamente, na íntegra, e
1042 avaliados quanto aos padrões da WWC (2017b). Primeiramente, os estudos
1043 foram separados em tipos de delineamento. Delineamentos intersujeitos foram
1044 avaliados em três etapas: tipo de delineamento, atrito da amostra e equivalência
1045 de linha de base entre os grupos. Delineamentos intrasujeitos foram avaliados
1046 usando os critérios piloto da WWC para delineamentos de sujeito-único (WWC,
1047 2017b, apêndice A), bem como os procedimentos de análise visual.

1048

1049 *3.3.1 Delineamentos intersujeitos*

1050

1051 Delineamentos intersujeitos foram classificados quanto ao método de
1052 distribuição dos participantes entre os grupos controle e experimental. Caso a
1053 distribuição fosse aleatorizada de alguma forma, o estudo era classificado como
1054 RCT, e, caso contrário, como QED.

1055 Estudos classificados como RCT passam pela avaliação do atrito da
1056 amostra, que consiste nas diferenças de mensuração dos resultados ocorridas
1057 devido a reduções na amostra inicial de participantes - ou seja, quando
1058 participantes que constavam inicialmente nos grupos abandonam a pesquisa ou

1059 quando seus resultados não podem ser utilizados devido a algum
1060 comprometimento (WWC, 2017b).

1061 Dois tipos de atrito foram calculados, o geral e o diferencial. O atrito geral
1062 consiste no percentual que a perda de participantes representa diante da
1063 amostra como um todo. O atrito diferencial consiste na diferença percentual da
1064 perda de participantes em relação aos grupos experimental e controle. Quando
1065 a combinação desses dois tipos de atrito é considerada alta, de acordo com os
1066 critérios fornecidos pelo manual (WWC, 2017b), então considera-se que o
1067 estudo possui viés por atrito alto, e o estudo não pode ser avaliado como
1068 "Elegível para atender aos padrões da WWC para Delineamentos de Grupo Sem
1069 Reservas" (a melhor avaliação atribuída pela instituição). Por outro lado, perdas
1070 de amostra que são resultantes de atos da natureza (como catástrofes naturais)
1071 ou devido a critérios estipulados antes da introdução da apresentação ou devido
1072 à aleatorização de uma subamostra em algum dos grupos não são consideradas
1073 como atrito.

1074 Pesquisas com atrito alto e todos os delineamentos classificados como
1075 QED são avaliadas quanto à equivalência de linha de base, que consiste no
1076 procedimento por meio do qual se avalia as diferenças entre as características
1077 dos grupos controle e experimental no pré-teste, ou seja, antes da administração
1078 da intervenção. O objetivo dessa etapa é verificar se os grupos possuem
1079 diferenças significativas entre si, antes da intervenção, ou se podem ser
1080 considerados equivalentes. O cálculo é realizado por meio da fórmula de
1081 tamanho de efeito, utilizando-se os dados do pré-teste de ambos os grupos².

1082 Caso os resultados apontem diferenças na característica avaliada
1083 maiores que 0.25 de desvio padrão, considera-se que os grupos não são
1084 equivalentes, e as pesquisas são avaliadas como "Não atendem aos padrões da
1085 WWC para Delineamentos de Grupo". Caso as diferenças estejam entre 0.05 e
1086 0.25 de desvio padrão, então a análise dos resultados da pesquisa deve
1087 apresentar os devidos ajustes estatísticos para satisfazer o critério de
1088 equivalência de linha de base. Caso a diferença seja menor que 0.05 de desvio
1089 padrão, então a pesquisa atende ao critério de equivalência de linha de base. Os
1090 RCTs com alto atrito e os QED's com equivalência de linha de base estabelecida

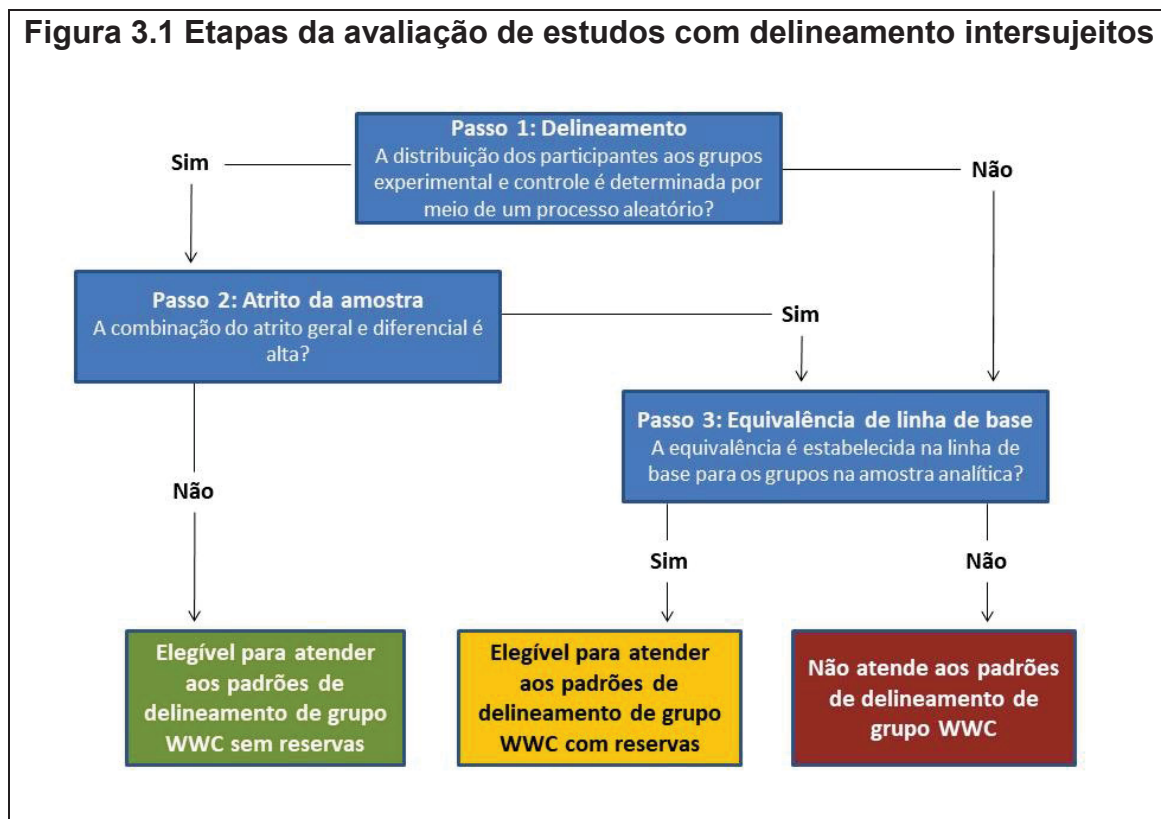
² Consultar WWC (2017a), apêndice E, para mais detalhes quanto ao cálculo de tamanho de efeito.

1091 são avaliados como Elegíveis para Atender aos Padrões da WWC para
1092 Delineamentos de Grupo com Reservas.

1093 A Figura 3.1 abaixo apresenta um diagrama com o fluxo das etapas da
1094 avaliação dos delineamentos intersujeitos que facilita o entendimento da
1095 dinâmica do processo.

1096

1097 **Figura 3.1 Etapas da avaliação de estudos com delineamento intersujeitos**



1098

1099

1100 3.3.2 Delineamentos intrasujeitos

1101

1102 Estudos com delineamento intrasujeito devem atender os critérios de
1103 elegibilidade para ter seus resultados avaliados e considerados na revisão
1104 (WWC, 2017b). Estudos que sejam classificados como "Atendem aos padrões-
1105 piloto da WWC para SCD Sem Reservas", ou "Atendem aos padrões-piloto da
1106 WWC para SCD com Reservas" têm seus resultados avaliados para ser
1107 classificados como fornecendo "Forte Evidência de uma Relação Causal",
1108 "Evidência Moderada de Relação Causal" ou "Nenhuma Evidência de Relação
1109 Causal". Estudos classificados como "Não atendem aos padrões-piloto da WWC
1110 para SCD" não são considerados para a avaliação do tipo de evidências que
1111 fornecem.

1112 Os critérios para avaliação de elegibilidade dos estudos SCD são (WWC,
1113 2017b):

- 1114
- 1115
- 1116
- 1117
- 1118
- 1119
- 1120
- 1121
- 1122
- 1123
- 1124
- 1125
- 1126
- 1127
- 1128
- 1129
- 1130
- 1131
- 1132
- 1133
- 1134
- 1135
- 1136
- 1137
- 1138
- 1139
- 1140
- 1141
- 1142
- 1143
- 1144
- A variável independente (a intervenção) deve ser sistematicamente manipulada, com o pesquisador determinando quando e como as condições da variável independente mudarão.
 - Para cada fase, a variável dependente deve ser mensurada sistematicamente pelo transcorrer do tempo por mais que um observador. O *Delineamento* precisa coletar concordância *interassessor*³ em cada fase de no mínimo 20% dos pontos de dados em cada condição (linha de base, intervenção) e a concordância *interassessor* deve atingir critérios mínimos: Concordância *interassessor* (comumente chamada de concordância interobservadores) deve ser documentada sobre a base de uma mensuração estatística da consistência do assessor. Ainda que haja mais do que 20 medidas estatísticas para representar concordância *interassessor* (e.g. Berk, 1979; Suen & Ary, 1989), medidas utilizadas comumente incluem concordância percentual (ou concordância proporcional) e coeficiente kappa de Cohen (Hartmann, Barrios, & Wood, 2004). De acordo com Hartmann et al. (2004), valores mínimos aceitáveis da concordância *interassessor* variam de 0.80 a 0.90 (na média) se medidos por concordância percentual e no mínimo 0.60 se mensurados pelo coeficiente kappa de Cohen.
 - O estudo deve incluir no mínimo três tentativas de demonstrar um efeito da intervenção em três pontos diferentes do tempo. O critério das três demonstrações é baseado em convenção profissional (Horner, Swaminathan, Sugai, & Smolkowski, 2012). Mais demonstrações posteriores aumentam a confiança no controle experimental (Kratochwill & Levin, 2010).
 - As fases devem atingir critérios envolvendo o número de pontos de dados para qualificar como uma tentativa de demonstrar um efeito.
 - *Reversão/remoção (AB)*: deve possuir um mínimo de quatro fases por caso com ao menos cinco pontos de dados⁴ por fase

³ Este termo não possui tradução direta (*interassessor*) ou equivalente para o português, uma vez o equivalente mais próximo, *interobservador*, é justamente o que os autores pretendem evitar, com o comentário em parênteses.

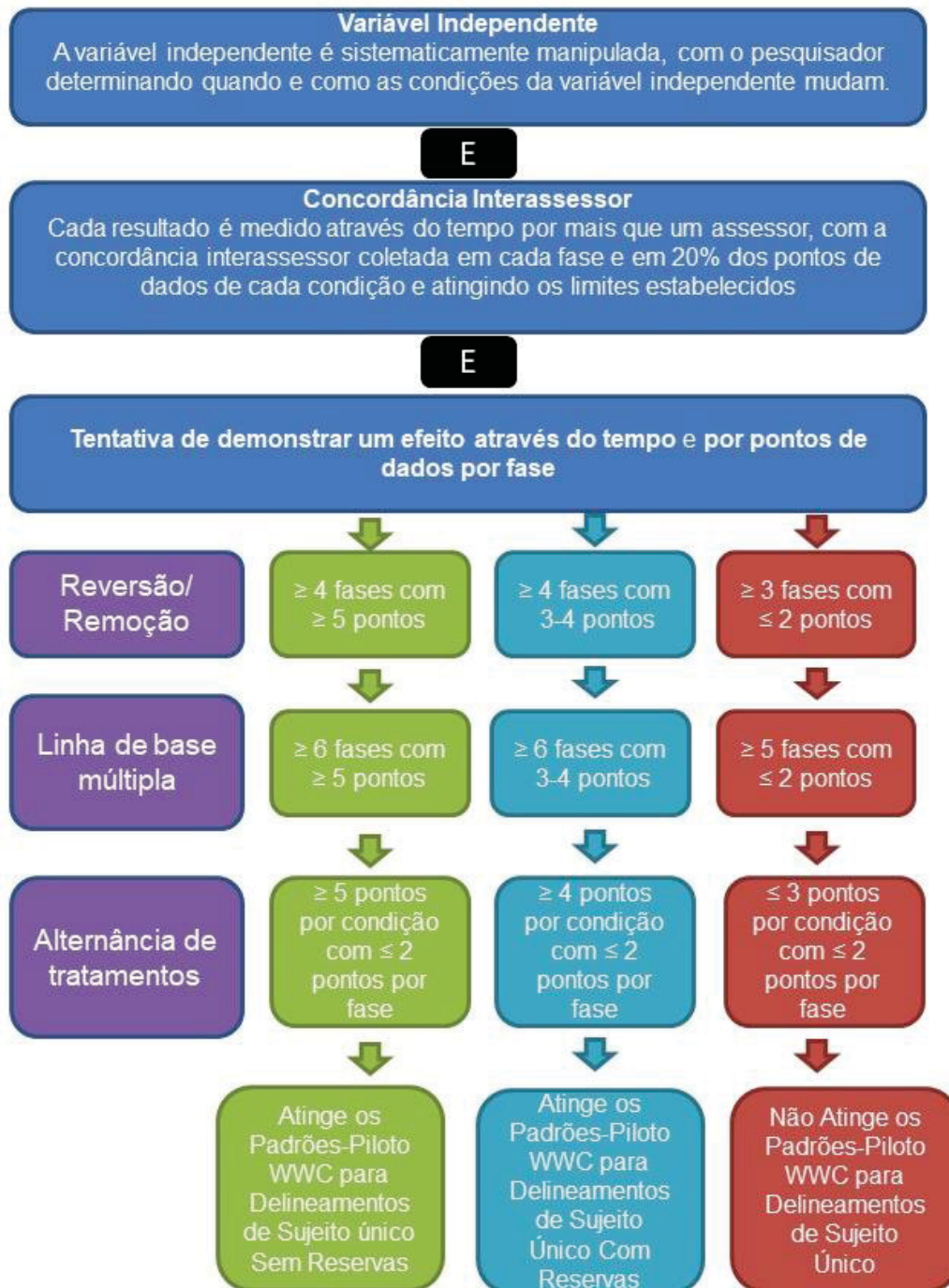
⁴ Um caso é definido, pela WWC (2017b), como "A unidade da administração da intervenção e análise de dados. Um caso pode ser um único participante ou um agrupamento de participantes (sala de aula ou comunidade)" (p. A2). Uma fase é definida como as diferentes condições nas quais a variável dependente é mensurada repetidamente (WWC, 2017b, p. A2). A expressão "pontos de dados" foi traduzida diretamente do original, *data points*. Poderíamos ter utilizado simplesmente a palavra "dados" como tradução, mas optamos por "pontos de dados" devido ao

- 1145 para *Atingir Sem Reservas os Padrões Piloto da WWC para*
 1146 *Delineamento de Sujeito Único*. Deve possuir no mínimo
 1147 quatro fases por caso com no mínimo três pontos de dados por
 1148 fase para *Atingir Com Reservas os Padrões Piloto da WWC*
 1149 *para Delineamento de Sujeito Único*. Quaisquer fases
 1150 baseadas em menos que três pontos de dados não podem ser
 1151 utilizadas para demonstrar a existência ou ausência de um
 1152 efeito.
- 1153 ○ *Linha de Base Múltipla e multiple probe*. Deve possuir no
 1154 mínimo seis fases com ao menos cinco pontos de dados por
 1155 fase para *Atingir Sem Reservas os Padrões Piloto da WWC*
 1156 *para Delineamento de Sujeito Único*. Deve possuir no mínimo
 1157 seis fases com no mínimo três pontos de dados por fase para
 1158 *Atingir Com Reservas os Padrões Piloto da WWC para*
 1159 *Delineamento de Sujeito Único*. Quaisquer fases baseadas em
 1160 menos que três pontos de dados não podem ser usadas para
 1161 demonstrar a existência ou ausência de um efeito. Ambos os
 1162 *Delineamentos* implicitamente requerem algum grau de
 1163 concorrência no tempo da sua implementação entre os casos
 1164 quando a intervenção está sendo introduzida. De outra forma,
 1165 esses *Delineamentos* não podem ser distinguidos de uma série
 1166 de *Delineamentos* AB separados.
 - 1167 ○ *Alternância de tratamentos*. Deve possuir no mínimo cinco
 1168 pontos de dados por condição (linha de base, intervenção) e
 1169 no máximo dois pontos de dados por fase para *Atingir Sem*
 1170 *Reservas os Padrões Piloto da WWC para Delineamento de*
 1171 *Sujeito Único*. Deve possuir quatro pontos de dados por
 1172 condição e no máximo dois pontos de dados por fase para
 1173 *Atingir com Reservas os Padrões Piloto da WWC para*
 1174 *Delineamento de Sujeito Único*. Quaisquer fases baseadas em
 1175 mais do que dois pontos de dados não pode ser usada para
 1176 demonstrar a existência ou ausência de um efeito porque esse
 1177 *Delineamento* tem como característica a rápida alternância
 1178 entre fases. Quando *Delineamentos* incluem comparações
 1179 entre múltiplas intervenções (por exemplo, A versus B, A
 1180 versus C, C versus B), cada comparação entre intervenções é
 1181 avaliada separadamente. (WWC, 2017, p. A3-5. Itálicos do
 1182 original.)

fato de que esse termo pode significar os dados representados como pontos nos gráficos típicos de delineamentos intrasujeitos.

1183 O fluxograma apresentado na Figura 3.2 a seguir, adaptado de WWC
 1184 (2017b), sintetiza os critérios para avaliação da elegibilidade de estudos SCD.
 1185

1186 **Figura 3.2 Critérios para avaliação da elegibilidade de estudos SCD**



1187

1188

1189 Estudos que forem classificados como "Atendem os padrões-piloto da
 1190 WWC para SCD" (com ou sem reservas) são avaliados a partir das etapas dos
 1191 procedimentos de análise visual descrita a seguir (WWC, 2017b). A primeira
 1192 etapa consiste na constatação de um padrão previsível nos dados da linha de
 1193 base, e tem por objetivo demonstrar que o problema ao qual a pesquisa se
 1194 dedica ocorre em certa frequência, além de fornecer dados suficientes para
 1195 demonstrar um padrão de responder claramente definido que sirva para avaliar
 1196 os efeitos da intervenção.

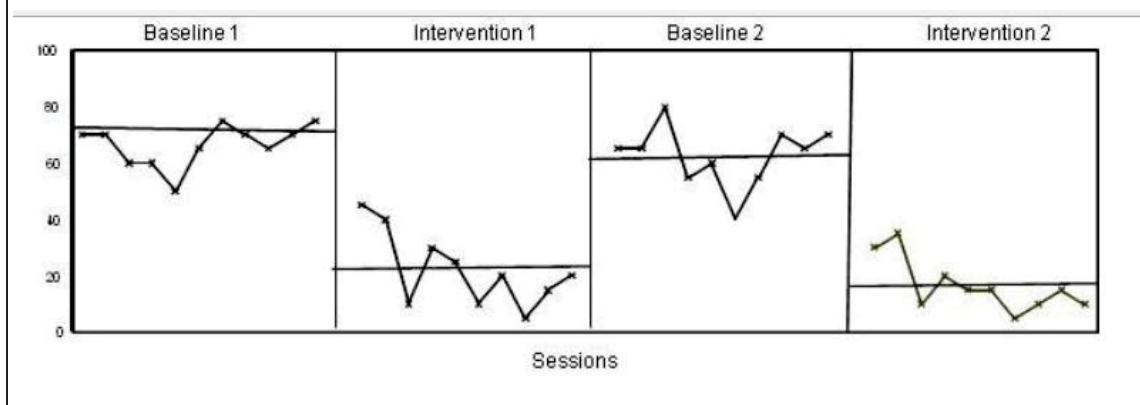
1197 Para a segunda etapa, a WWC (2017b) propõe critérios para se analisar
 1198 cada fase do experimento (análise intra-fase), conforme segue:

- 1199 ○ *Nível* se refere à média do escore para os dados dentro de uma
- 1200 fase.
- 1201 ○ *Tendência* se refere à inclinação da linha reta que melhor se
- 1202 encaixa nos dados dentro de uma fase.
- 1203 ○ *Variabilidade* se refere à abrangência ou desvio padrão dos dados
- 1204 sobre a linha reta que melhor se encaixa. (WWC, 2017b, p. A7.
- 1205 Itálicos do original.)
- 1206

1207 Na Figura 3.3 abaixo, adaptada de WWC (2017b), é possível observar
 1208 um exemplo da análise do nível dos pontos de dados em um delineamento
 1209 ABAB.

1210

1211 **Figura 3.3 Exemplo de análise do nível dos pontos de dados em**
 1212 **delineamento ABAB**



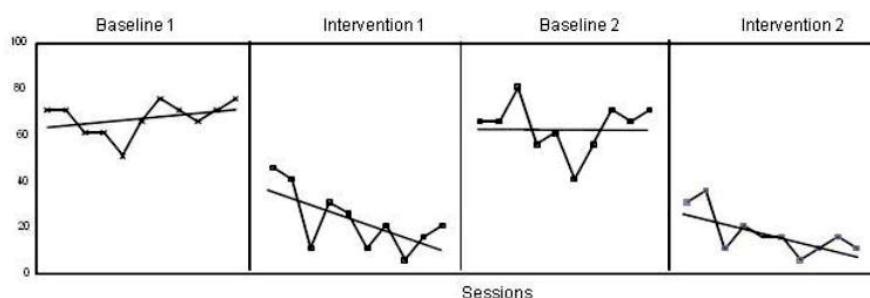
1213

1214

1215 Na próxima Figura 3.4, adaptada de WWC (2017b), é possível observar
 1216 um exemplo da análise da tendência dos pontos de dados em um delineamento
 1217 ABAB.

1218

1219 **Figura 3.4 Exemplo de análise da tendência dos pontos de dados em**
 1220 **delineamento ABAB**



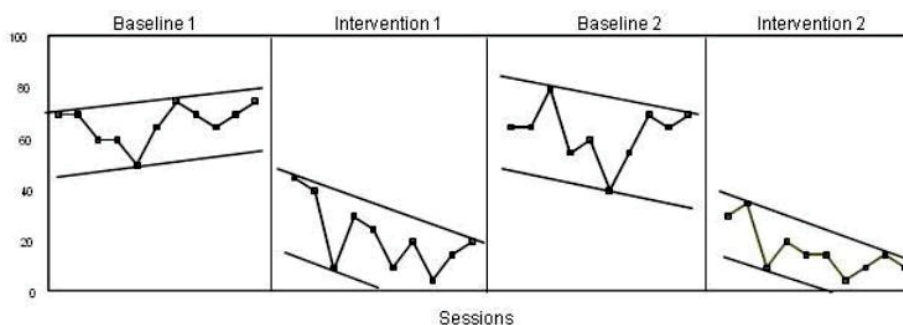
1221

1222

1223 Na Figura 3.5 seguinte, adaptada de WWC (2017b), é possível observar
 1224 um exemplo da análise da variabilidade dos pontos de dados em um
 1225 delineamento ABAB.

1226

1227 **Figura 3.5 Exemplo de análise da variabilidade dos pontos de dados em**
 1228 **delineamento ABAB**



1229

1230

1231 A análise proporcionada pela segunda etapa deve ser complementada,
 1232 de acordo com a WWC (2017b), com a avaliação proposta na terceira etapa, que
 1233 consiste na comparação entre os dados de cada fase (avaliação interfases). Mais
 1234 uma vez, três critérios são propostos para a condução dessa avaliação:

1235

1236

1237

1238

- *Imediatismo do efeito* refere-se à mudança no nível entre os últimos três pontos de dados coletados e os primeiros três pontos de dados seguintes. Quanto mais rápida (ou mais imediata), mais convincente é a inferência de que a mudança

1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259

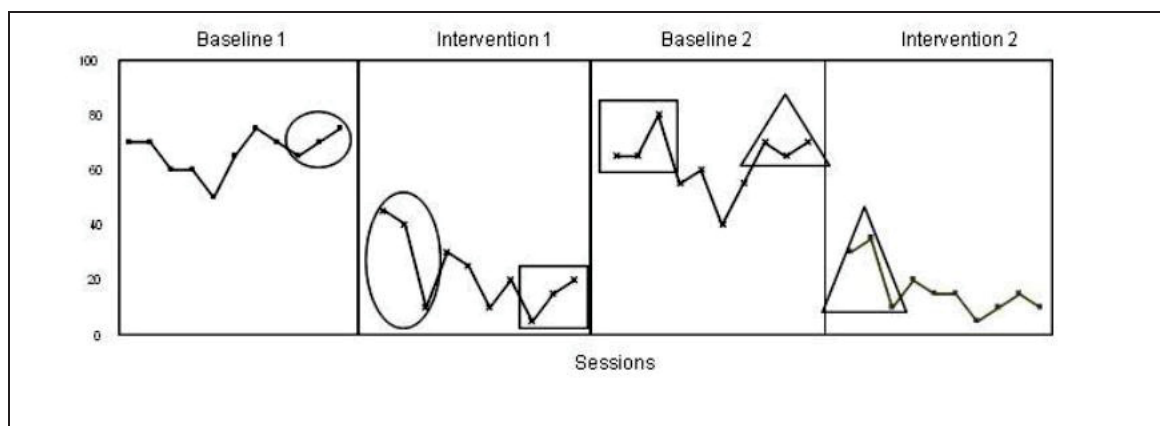
na medida da variável dependente ocorreu devido à manipulação da variável independente. Efeitos atrasados podem comprometer a validade interna do *Delineamento*. Todavia, atrasos previstos no efeito ou efeitos graduais da intervenção podem ser planejados no *Delineamento* de maneira a influenciar decisões sobre duração das fases em um estudo particular.

- *Sobreposição* refere-se à proporção dos dados de uma fase que se sobrepõem com os dados da fase anterior. Quanto menor a proporção dos pontos de dados que se sobrepõem (ou inversamente, quanto maior a separação), mais convincente a demonstração de um efeito.
- *Consistência dos dados em fases similares* envolve olhar para os dados de todas as fases dentro das mesmas condições (por exemplo, todas as fases de linha de base; todas as fases “atividade em pares”) e examinar a extensão com que há consistência nos padrões dos dados das fases com as mesmas condições. Quanto maior a consistência, maior a probabilidade de que os dados representem uma relação causal. (WWC, 2017b, p.A7. Itálico no original.)

1260 Na próxima Figura 3.6, adaptada de WWC (2017b), é possível observar
1261 um exemplo da análise do imediatismo do efeito da intervenção na transição de
1262 cada fase, em um delineamento ABAB.

1263

1264 **Figura 3.6 Exemplo de análise do imediatismo do efeito da intervenção em**
1265 **cada fase**



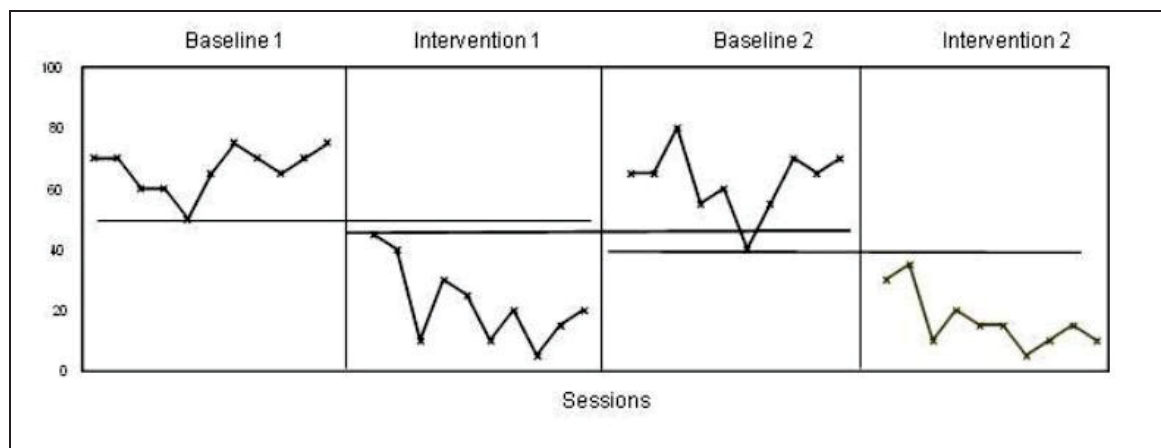
1266

1267

1268 Na Figura 3.7 abaixo, adaptada de WWC (2017b), é possível observar
 1269 um exemplo da análise da sobreposição entre as fases de um delineamento
 1270 ABAB.

1271

1272 **Figura 3.7 Exemplo de análise de sobreposição entre as fases**



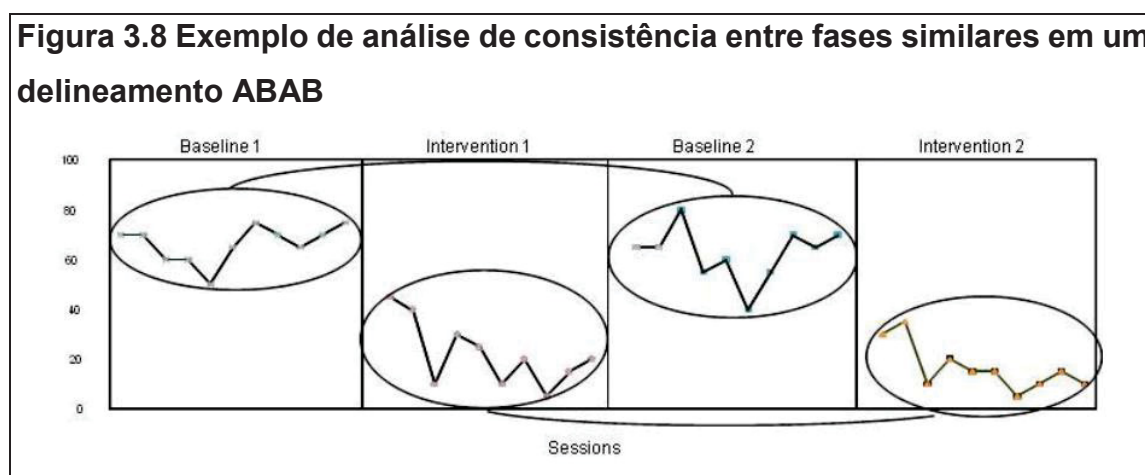
1273

1274

1275 Na Figura 3.8 seguinte, adaptada de WWC (2017b), é possível observar
 1276 um exemplo da análise de consistência entre fases similares em um
 1277 delineamento ABAB.

1278

1279 **Figura 3.8 Exemplo de análise de consistência entre fases similares em um**
 1280 **delineamento ABAB**



1281

1282

1283 A quarta etapa deste procedimento consiste em comparar as
 1284 informações coletadas da análise e avaliação entre as fases para determinar se
 1285 os dados da pesquisa sob avaliação atingem os critérios de demonstração de
 1286 um efeito em três pontos temporais diferentes.

1287 A partir de tais etapas de análise, a WWC (2017b) propõe três
 1288 classificações para os estudos revisados. Se uma relação causal é demonstrada

1289 no estudo, então ele apresenta “Forte Evidência de uma Relação Causal”. Isso
1290 é constatado quando o estudo apresenta no mínimo três demonstrações do
1291 efeito da intervenção, tal como avaliado nos critérios das quatro etapas
1292 anteriormente descritas. Para ser classificado como apresentando “Forte
1293 Evidência”, o estudo não deve apresentar casos de “nenhum efeito”.

1294 Se a pesquisa revisada apresenta três demonstrações da ocorrência do
1295 efeito da intervenção, e também apresenta no mínimo um caso em que nenhum
1296 efeito ocorreu, então ela é classificada como apresentando “Evidência Moderada
1297 de uma Relação Causal” (What Works Clearinghouse, 2017b). Se a pesquisa
1298 revisada não apresenta pelo menos três demonstrações da ocorrência do efeito,
1299 segundo os critérios apresentados anteriormente, então ela é classificada como
1300 “Nenhuma Evidência de uma Relação Causal”.

1301 Os seguintes tópicos devem ser considerados para identificar um caso
1302 de nenhum efeito:

- 1303 • Dados dentro da fase de linha de base não proveem demonstração
1304 suficiente de um padrão claramente definido de responder que
1305 possa ser usado para projetar o desempenho esperado assumindo
1306 que nenhuma mudança ocorreu devido à não manipulação da
1307 variável independente.
- 1308 • Falha em estabelecer um padrão consistente dentro de cada fase
1309 (por exemplo, alta variabilidade dentro de uma fase).
- 1310 • (A) Grande latência entre a introdução da variável independente e
1311 a mudança na variável dependente ou (B) sobreposição entre
1312 padrões observados e projetados da variável dependente entre as
1313 fases de linha de base e de intervenção, tornando difícil determinar
1314 se a intervenção foi responsável pelo efeito desejado.
- 1315 • Padrões inconsistentes entre fases semelhantes (por exemplo, um
1316 Delineamento ABAB no qual quando a primeira fase de intervenção
1317 aconteceu, a variável dependente apresentou alta frequência nos
1318 dados; e na segunda vez que a intervenção foi introduzida, a
1319 variável independente apresentou baixa frequência, e assim
1320 sucessivamente).
- 1321 • Comparar os padrões projetados e observados da variável
1322 dependente entre as fases não demonstra evidência de uma
1323 relação causal (por exemplo, não há no mínimo três
1324 demonstrações do efeito da intervenção). (WWC, 2017b, p. A14)

1325

1326 Por fim, a WWC (2017b) faz uma observação adicional em relação aos
1327 experimentos com linha de base múltipla:

1328 Ao examinar um *Delineamento* de linha de base múltipla,
1329 considere também em que medida o tempo no qual um efeito básico é
1330 inicialmente demonstrado com uma série (por exemplo, nos primeiros
1331 cinco dias após a introdução da intervenção para o Participante #1)
1332 está associado com mudanças nos padrões de dados durante o
1333 mesmo período de tempo nas outras séries do *Delineamento* (por
1334 exemplo, os mesmos cinco dias para os Participantes #2, #3, e #4). Se
1335 o efeito básico é demonstrado em uma série, e há mudança nos
1336 padrões de dados em outras séries, a maior avaliação que o
1337 *Delineamento* pode receber é *Evidência Moderada* (WWC, 2017b, p.
1338 A14. itálico no original)

1339
1340

1341 3.3.3 Síntese dos resultados

1342

1343 Ao fim da revisão, a WWC (2017a) recomenda a integração dos
1344 resultados de estudos com delineamentos intersujeitos em sumários
1345 quantitativos e qualitativos. Quantitativamente, são calculadas a magnitude dos
1346 resultados e a significância estatística deles. Quanto à magnitude, a WWC
1347 (2017a) fornece diretrizes para se calcular o tamanho de efeito dos resultados,
1348 que consiste na diferença da média do desempenho entre grupo controle e
1349 experimental, dividida pelo desvio padrão dos dados agregados entre os grupos
1350 (no original, "the pooled within-group standard deviation", WWC, 2017a, p. E2).

1351 Quanto à significância estatística, a principal medida consiste no índice
1352 de aprimoramento, que consiste na diferença entre o ranque percentual
1353 correspondente à média do valor dos resultados para o grupo de intervenção e
1354 o grupo controle, ou seja, a mudança que se espera obter na pontuação
1355 percentual de um estudante mediano no grupo controle ao receber a intervenção.
1356 Além dessa medida, a significância estatística deve ser calculada como forma
1357 de estimar a probabilidade de se observar o efeito estudado, a qual deve ser
1358 maior que um em vinte (WWC, 2017a).

1359 Como forma de agregar qualitativamente os resultados, a WWC (2017a)
1360 propõe que sejam apresentados sumários que indiquem a direção e a força dos
1361 resultados. A partir das estimativas de tamanho de efeito e significância

1362 estatística, os resultados de cada estudo devem ser classificados em uma das
1363 cinco categorias apresentadas na Tabela 1 a seguir, adaptada de WWC (2017a):

1364

1365 Tabela 1

1366 *Caracterização dos resultados de um efeito baseado em medida individual dentro de*
1367 *um domínio*

Efeito positivo estatisticamente significativo	O efeito estimado é positivo ⁵ e estatisticamente significativo
Efeito positivo substancialmente importante	O efeito estimado é positivo, estatisticamente não é significativo mas é substancialmente importante
Efeito indeterminado	O efeito estimado não é nem estatisticamente significativo nem substancialmente importante
Efeito negativo substancialmente importante	O efeito estimado é negativo, estatisticamente não é significativo mas é substancialmente importante
Efeito negativo estatisticamente significativo	O efeito estimado é negativo e estatisticamente significativo

Nota: Uma estimativa estatisticamente significativa de um efeito é uma na qual a probabilidade de se observar um efeito que seja ao menos tão grande quanto o efeito mensurado sob a perspectiva de que a intervenção não teve impacto é menor de que um em 20 (usando um teste-t, bicaudal com $p = 0,05$). Um tamanho de efeito de 0,25 desvios padrões ou maior é considerado substancialmente importante.

1368

1369 Um relatório acerca da intervenção sendo revisada deve ser
1370 apresentado com uma classificação da intervenção e a extensão das evidências
1371 para essa classificação. A Tabela 2 abaixo, adaptada de WWC (2017a),
1372 apresenta os critérios propostos para avaliação da efetividade de uma
1373 intervenção. Vale lembrar que esses critérios não são excludentes, e uma
1374 intervenção que atinja múltiplos critérios deve receber a maior classificação
1375 atingida.

1376

⁵ Efeitos positivos e negativos podem ser considerados como direções do efeito mensurado. Dessa forma, um efeito positivo é um acréscimo no desempenho avaliado, e um efeito negativo, uma redução do desempenho.

1377

1378 Tabela 2

1379 *Critérios usados para determinar a efetividade de uma intervenção*

Efeitos positivos: evidência forte de um efeito positivo com nenhuma evidência contrária que a supere	<ul style="list-style-type: none"> • Dois ou mais estudos que mostram efeitos positivos estatisticamente significantes com ao menos um dos quais Atende aos Padrões WWC de Delineamento de Grupo Sem Reservas, E • Nenhum estudo que apresenta efeitos negativos estatisticamente significantes ou substancialmente importantes.
Efeitos potencialmente positivos: evidências de um efeito positivo com nenhuma evidência contrária que a supere	<ul style="list-style-type: none"> • Ao menos um estudo que mostra efeitos positivos estatisticamente significantes ou substancialmente importantes, E • Uma quantidade menor ou igual de estudos que mostram efeitos indeterminados do que estudos que mostram efeitos positivos estatisticamente significantes ou substancialmente importantes • Nenhum estudo que mostra efeitos negativos estatisticamente significantes ou substancialmente importantes
Efeitos indiscerníveis: nenhuma evidência afirmativa de efeitos	<ul style="list-style-type: none"> • Nenhum estudo que mostre efeitos estatisticamente significativos ou substancialmente importantes, sejam negativos ou positivos
Efeitos mistos: evidência de efeitos inconsistentes	<p>AMBOS os seguintes:</p> <ul style="list-style-type: none"> • Ao menos um estudo que mostra efeitos positivos estatisticamente significantes ou substancialmente importantes, E • Ao menos um estudo que mostra efeitos negativos estatisticamente significantes ou substancialmente

importantes, MAS não mais do que o número de estudos que mostram efeitos positivos estatisticamente significantes ou substancialmente importantes

OU ambos os seguintes:

- Ao menos um estudo que mostra efeitos estatisticamente importantes ou substancialmente importantes, E
- Mais estudos que mostram efeitos indeterminados do que estudos que mostram efeitos estatisticamente importantes ou substancialmente importantes

Efeitos potencialmente negativos: evidência de um efeito negativo com nenhuma evidência contrária que a supere

AMBOS os seguintes:

- Ao menos um estudo que mostra efeitos negativos estatisticamente significantes ou substancialmente importantes, E
- Nenhum estudo que mostre efeitos positivos estatisticamente significantes ou substancialmente relevantes;

OU ambos os seguintes:

- Dois ou mais estudos que mostrem efeitos negativos estatisticamente significantes ou substancialmente importantes, ao menos um estudo que mostre efeitos positivos estatisticamente significantes ou substancialmente importantes, E
 - Mais estudos mostrando efeitos negativos estatisticamente significantes ou substancialmente importante do que mostrando efeitos positivos estatisticamente significantes ou substancialmente importantes
-

Efeitos negativos: forte evidência de um efeito negativo com nenhuma evidência contrária que a supere	<ul style="list-style-type: none"> • Dois ou mais estudos que mostram efeitos negativos estatisticamente significantes com ao menos um dos quais Atende aos Padrões WWC de Delineamento de Grupo Sem Reservas, E • Nenhum estudo que apresenta efeitos negativos estatisticamente significantes ou substancialmente importantes.
---	--

1380

1381 Para se caracterizar a extensão dos resultados de uma intervenção, a
 1382 WWC (2017a) recomenda a combinação dos resultados entre os estudos
 1383 revisados como forma de determinar a classificação da intervenção, de maneira
 1384 que seja possível informar aos leitores sobre quanta evidência foi utilizada para
 1385 se determinar a classificação da intervenção. A Tabela 3 a seguir, adaptada de
 1386 WWC (2017a), apresenta os critérios propostos para determinar a extensão das
 1387 evidências de uma intervenção.

1388

1389 Tabela 3

1390 *Critérios usados para determinar a extensão das evidências de uma intervenção*

De média a ampla	<ul style="list-style-type: none"> • O domínio inclui mais que um estudo, E • O domínio inclui mais que um cenário (<i>setting</i>), E • O domínio dos resultados é baseado em uma amostra total de, no mínimo, 350 estudantes, OU, assumindo 25 estudantes por sala de aula, um total de no mínimo 14 salas de aula entre os estudos
Pequena	<ul style="list-style-type: none"> • O domínio inclui apenas um estudo, OU • O domínio inclui apenas um cenário (<i>setting</i>), OU • O domínio dos resultados é baseado em uma amostra total de menos que 350 estudantes, OU, assumindo 25

estudantes por sala de aula, um total de menos que 14 salas de aula entre os estudos

1391

1392 Por fim, a WWC (2017a) elabora guias práticos que integram as
1393 informações sobre os resultados da revisão realizada. Esses resultados
1394 apresentam as seguintes informações: o número de estudos, a qualidade dos
1395 estudos, se os estudos representam a gama (*range*) de participantes, contextos
1396 e comparações nas quais a recomendação é focada, se os resultados dos
1397 estudos podem ser atribuídos às práticas recomendadas e se os resultados dos
1398 estudos são consistentemente positivos.

1399 Quanto aos resultados de estudos que utilizam delineamentos
1400 intrasujeitos, a WWC (2017b) fornece critérios para elaboração da classificação
1401 de uma intervenção com base na análise visual, uma vez que métodos para o
1402 cálculo de tamanho de efeito utilizando SCD ainda não foram desenvolvidos.
1403 Dessa forma, emprega-se os critérios a seguir para integrar em uma síntese
1404 classificativa os estudos revisados sobre uma intervenção:

- 1405
- 1406 • No mínimo cinco estudos de sujeito único apresentam resultados
1407 que Atingem Sem Reservas os Padrões WWC ou Atingem com
Reservas os Padrões WWC .
 - 1408 • Os estudos de caso único devem ter sido conduzidos por no
1409 mínimo três equipes de pesquisa diferentes sem sobreposição de
1410 autores entre três instituições.
 - 1411 • O número de casos únicos nos Delineamentos, quando somados,
1412 devem atingir um mínimo de 20.

1413 Os procedimentos apresentados pela WWC (2017b) não apresentam
1414 mais detalhes acerca da agregação dos resultados de estudos SCD em sínteses
1415 qualitativas, e, dessa forma, não fica claro como avaliar a direção do efeito e a
1416 extensão de suas evidências, uma vez que os critérios utilizados requerem o
1417 cálculo do tamanho de efeito e da significância estatística da intervenção.

1418 4 RESULTADOS

1419

1420

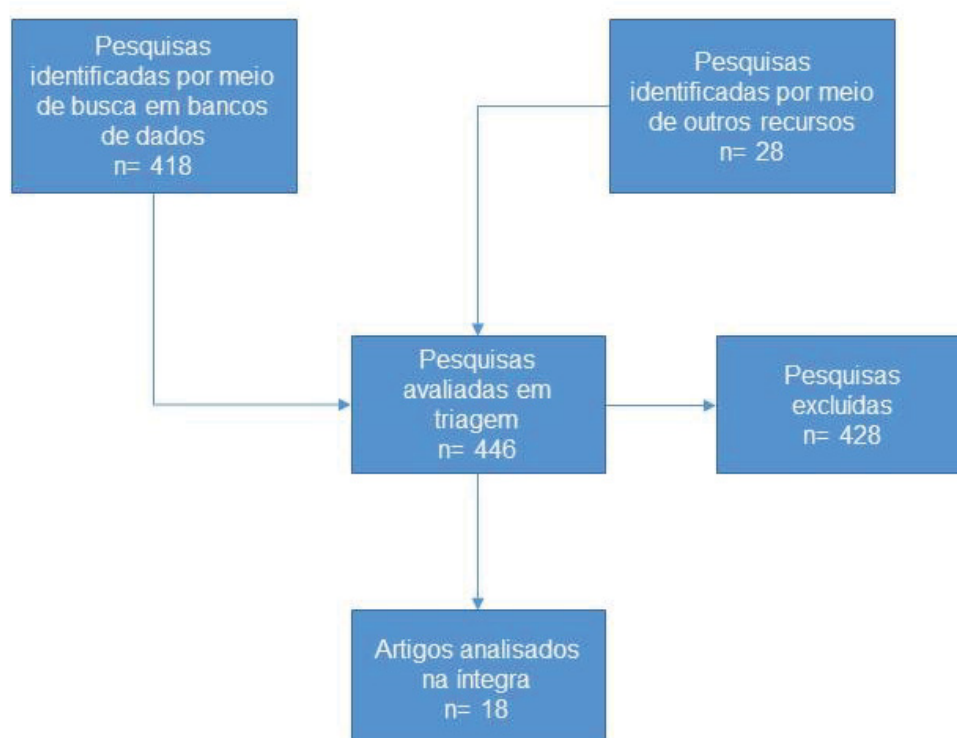
1421 **4.1 Artigos triados**

1422

1423 Foram encontradas 418 pesquisas por meio da busca inicial nos bancos
1424 de dados. Com a eliminação das pesquisas duplicadas, e com a triagem inicial,
1425 14 artigos constituíram os resultados preliminares. Após a leitura desses artigos,
1426 28 outros foram buscados por meio de referências que possivelmente se
1427 enquadravam nos critérios de inclusão desta revisão. A triagem resultante
1428 desses artigos levou à identificação de mais 4 pesquisas, constituindo, portanto,
1429 18 artigos como resultado final para avaliação, como consta na Figura 4.1
1430 abaixo.

1431

Figura 4.1. Fluxograma dos artigos encontrados para revisão



1432

1433 A seguir é apresentada a lista completa dos artigos resultantes da
1434 triagem:

- 1435 • Fields, L., Travis, R., Roy, D., Yadlovker, E., Aguiar-Rocha, L. D.,
1436 & Sturmey, P. (2009). Equivalence class formation: A method for
1437 teaching statistical interactions. *Journal of Applied Behavior*
1438 *Analysis*, 42(3), 575-593.
- 1439 • Fienup, D. M., Critchfield, T. S., & Covey, D. P. (2009). Building
1440 contextually-controlled equivalence classes to teach about
1441 inferential statistics: A preliminary demonstration. *Experimental*
1442 *Analysis of Human Behavior Bulletin*, 27, 1-10.
- 1443 • Critchfield, T. S., & Fienup, D. M. (2010). Using stimulus
1444 equivalence technology to teach statistical inference in a group
1445 setting. *Journal of Applied Behavior Analysis*, 43(4), 763-768.
- 1446 • Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching
1447 brain—behavior relations economically with stimulus equivalence
1448 technology. *Journal of Applied Behavior Analysis*, 43(1), 19-33.
- 1449 • Walker, B. D., Rehfeldt, R. A., & Ninness, C. (2010). Using the
1450 stimulus equivalence paradigm to teach course material in an
1451 undergraduate rehabilitation course. *Journal of Applied Behavior*
1452 *Analysis*, 43(4), 615-633.
- 1453 • Fienup, D. M., & Critchfield, T. S. (2011). Transportability of
1454 equivalence-based programmed instruction: Efficacy and
1455 efficiency in a college classroom. *Journal of Applied Behavior*
1456 *Analysis*, 44(3), 435-450.
- 1457 • Lovett, S., Rehfeldt, R. A., Garcia, Y., & Dunning, J. (2011).
1458 Comparison of a stimulus equivalence protocol and traditional
1459 lecture for teaching single-subject designs. *Journal of Applied*
1460 *Behavior Analysis*, 44(4), 819-833.
- 1461 • Zinn, T. E., Newland, M. C., & Ritchie, K. E. (2015). The efficiency
1462 and efficacy of equivalence-based learning: A randomized
1463 controlled trial. *Journal of Applied Behavior Analysis*, 48(4), 865-
1464 882.

- 1465 • Pytte, C. L., & Fienup, D. M. (2012). Using equivalence-based
1466 instruction to increase efficiency in teaching
1467 neuroanatomy. *Journal of Undergraduate Neuroscience*
1468 *Education*, 10(2), A125.
- 1469 • Walker, B. D., & Rehfeldt, R. A. (2012). An evaluation of the
1470 stimulus equivalence paradigm to teach single-subject design to
1471 distance education students via blackboard. *Journal of Applied*
1472 *Behavior Analysis*, 45(2), 329-344.
- 1473 • Critchfield, T. S. (2014). Online equivalence-based instruction
1474 about statistical inference using written explanation instead of
1475 match-to-sample training. *Journal of Applied Behavior*
1476 *Analysis*, 47(3), 606-611.
- 1477 • Albright, L., Reeve, K. F., Reeve, S. A., & Kisamore, A. N. (2015).
1478 Teaching statistical variability with equivalence-based
1479 instruction. *Journal of Applied Behavior Analysis*, 48(4), 883-894.
- 1480 • Fienup, D. M., Wright, N. A., & Fields, L. (2015). Optimizing
1481 equivalence-based instruction: Effects of training protocols on
1482 equivalence class formation. *Journal of Applied Behavior*
1483 *Analysis*, 48(3), 613-631.
- 1484 • O'Neill, J., Rehfeldt, R. A., Ninness, C., Muñoz, B. E., & Mellor, J.
1485 (2015). Learning Skinner's verbal operants: Comparing an online
1486 stimulus equivalence procedure to an assigned reading. *The*
1487 *Analysis of Verbal Behavior*, 31(2), 255-266.
- 1488 • Reyes-Giordano, K., & Fienup, D. M. (2015). Emergence of
1489 topographical responding following equivalence-based
1490 neuroanatomy instruction. *The Psychological Record*, 65(3), 495-
1491 507.
- 1492 • Zinn, T. E., Newland, M. C., & Ritchie, K. E. (2015). The efficiency
1493 and efficacy of equivalence-based learning: A randomized
1494 controlled trial. *Journal of Applied Behavior Analysis*, 48(4), 865-
1495 882.
- 1496 • Albright, L., Schnell, L., Reeve, K. F., & Sidener, T. M. (2016).
1497 Using stimulus equivalence-based instruction to teach graduate

- 1498 students in applied behavior analysis to interpret operant functions
1499 of behavior. *Journal of Behavioral Education*, 25(3), 290-309.
- 1500 • Fienup, D. M., Mylan, S. E., Brodsky, J., & Pytte, C. (2016). From
1501 the laboratory to the classroom: The effects of equivalence-based
1502 instruction on neuroanatomy competencies. *Journal of Behavioral
1503 Education*, 25(2), 143-165.
- 1504 • Greville, W. J., Dymond, S., & Newton, P. M. (2016). The student
1505 experience of applied equivalence-based instruction for
1506 neuroanatomy teaching. *Journal of Educational Evaluation for
1507 Health Professions*, 13.

1508

1509 Os artigos que constituíram o material a ser avaliado nesta pesquisa
1510 foram resumidos a seguir, buscando sintetizar as seguintes informações:
1511 autores, ano de publicação, objetivos do estudo, comportamentos ensinados,
1512 delineamento utilizado e resumo geral⁶ dos resultados.

1513 Fields, Travis, Roy, Yadlovker, Aguiar-Rocha e Sturmey (2009)
1514 realizaram um dos primeiros estudos que utilizaram procedimentos de formação
1515 de classes de equivalência para ensinar repertórios acadêmicos em alunos do
1516 ensino superior. Nessa pesquisa, eles desenvolveram um procedimento para
1517 ensinar interações estatísticas a um grupo de estudantes, enquanto o outro
1518 grupo, que não recebia a instrução, servia de controle. Em um delineamento pré-
1519 teste/intervenção/pós-teste, no qual os alunos eram aleatoriamente designados
1520 a cada grupo, os pesquisadores descobriram que os estudantes no grupo
1521 experimental tiveram um desempenho que foi, em média, 35% maior do que
1522 aqueles do grupo controle.

1523 Fienup, Crithcfield e Covey (2009), também podem ser considerados
1524 pioneiros no uso de EBI para desenvolver repertórios acadêmicos em alunos do
1525 ensino superior. Em um estudo piloto, conduziram dois experimentos nos quais
1526 dois grupos de voluntários recebiam, ambos, as intervenções, em um

⁶ Ao apresentar sinteticamente os resultados das pesquisas tentamos inicialmente, por padrão, apresentar os dados quantitativos. Contudo, ao longo do processo percebemos que em vários artigos a complexidade desses dados demandaria explicações extensas, prejudicando a própria síntese. Algumas pesquisas, por exemplo, apresentaram apenas resultados particulares para cada conjunto de relações entre estímulos aprendidas por cada participante. Em vista disso, optou-se por apresentar os resultados com linguagem mais abrangente, embora eventualmente menos precisa.

1527 delineamento AB. No primeiro experimento, conceitos de estatística inferencial
1528 foram ensinados. No entanto, devido a um equívoco com um dos conceitos
1529 utilizados ao ensinar os comportamentos aos alunos, o segundo experimento
1530 visou corrigir esse problema. Em ambos os experimentos, os alunos
1531 aprenderam, com praticamente 100% de acerto, todas as relações entre
1532 estímulos ensinadas, mesmo as que não foram diretamente treinadas.

1533 Fienup e Critchfield (2010) utilizaram de procedimentos de formação de
1534 classes de equivalência entre estímulos ao ensinar a 10 graduandos repertórios
1535 de estatística inferencial por meio de um delineamento pré-
1536 teste/intervenção/pós-teste. Os participantes apresentaram um aumento médio
1537 no desempenho de 59% nos escores entre avaliação pré e pós-intervenção nos
1538 comportamentos relacionados à estatística inferencial, de 52% nos escores pré
1539 e pós-intervenção nos comportamentos relacionados à decisões baseadas em
1540 hipóteses científicas e direção de um efeito, e de 36% nos escores pré e pós-
1541 intervenção nos comportamentos relacionados à influência condicional da
1542 estatística inferencial sobre decisões a respeito de hipóteses científicas e nulas.

1543 Fienup, Covey e Critchfield (2010) estenderam o leque de áreas da
1544 aplicação da tecnologia de ensino baseado em equivalência para ensinar a
1545 quatro alunos de graduação relações entre classes de estímulos referentes
1546 neuroanatomia, por meio do uso de um delineamento intrasujeito conhecido
1547 como *multiple probe baseline*. Os alunos apresentaram um ganho médio de
1548 aproximadamente 50% entre pré-teste e pós-teste na quantidade de tentativas
1549 corretas durante um tempo de dedicação máximo de dezessete minutos, para
1550 aprender as relações entre dois conjuntos de classe de estímulos.

1551 Walker, Rehfeldt e Ninnes (2010) empregaram três protocolos para
1552 ensinar a alunos de uma disciplina sobre necessidades especiais repertórios
1553 relacionados a três temas do curso, utilizando EBI, mas de uma maneira que se
1554 assemelhasse mais ao contexto de uma sala de aula. Dois experimentos foram
1555 realizados, e, em ambos, atividades na forma de questionários em folhas de
1556 caderno (*worksheets*) foram utilizadas como forma de ensinar as relações entre
1557 estímulos, ao invés de programas de computador, como nas pesquisas
1558 anteriores. No primeiro experimento, os estudantes passaram por todos os
1559 protocolos de ensino e deveriam responder a testes pós-intervenção em folhas
1560 de papel no final do semestre, além de realizar testes pós-intervenção oralmente

1561 logo após a intervenção. No segundo experimento, como os participantes não
1562 eram estudantes da disciplina na qual foi realizada o experimento, eles fizeram
1563 o teste pós-intervenção escrito logo após a intervenção. Ambos os experimentos
1564 utilizaram um delineamento pré/pós-intervenção. Os resultados indicaram que a
1565 maioria dos participantes adquiriram os comportamentos ensinados, mesmo
1566 aqueles que emergiram por meio da formação de equivalência.

1567 Fienup e Critchfield (2011) replicaram sua pesquisa anterior (Fienup e
1568 Critchfield, 2010) ao avaliar o efeito do uso de instrução baseada em formação
1569 de classes de equivalência para ensinar a alunos de graduação conceitos sobre
1570 estatística inferencial e tomada de decisão baseada em hipóteses. Contudo,
1571 desta vez utilizaram dois grupos controle: um negativo, no qual os participantes
1572 não recebiam instrução alguma, e um grupo controle com instrução completa, no
1573 qual todas as relações eram ensinadas diretamente. O objetivo era avaliar se o
1574 grupo que recebeu instrução baseada em equivalência aprenderia todas as
1575 relações passando por menos treino direto. De acordo com Fienup e Critchfield
1576 (2011), foi observado um acréscimo no desempenho médio do grupo
1577 experimental entre a avaliação pré e pós teste de 49,3% de respostas corretas
1578 para 93,2% de respostas corretas, e também no grupo controle com instrução
1579 total, que aumentou o percentual de respostas corretas de 49,6% para 89,6%,
1580 enquanto que o mesmo crescimento não pode ser observado no grupo controle
1581 negativo, que obteve desempenho médio de 49,1% no pré teste e 50,7% no pós
1582 teste. Além disso, os tamanhos de efeito foram de 3,8 ao se comparar o
1583 desempenho do grupo experimental com o grupo controle negativo, e de 3,3
1584 entre o grupo de instrução total e o grupo controle. Os participantes do grupo de
1585 instrução completa levaram três vezes mais tempo para realizar o primeiro grupo
1586 de atividades do experimento que o grupo experimental, 1,6 vezes mais tempo
1587 na segunda lição, e conseguiram fazer a terceira lição com o mesmo tempo que
1588 o grupo experimental. Os resultados sugerem, de acordo com os autores, que o
1589 ensino de estatística baseado em equivalência não apenas é efetivo, como
1590 requer menos tempo de dedicação do que o ensino direto de todas as relações.

1591 Lovett, Rehfeldt, Garcia e Dunning (2011) compararam a efetividade do
1592 uso de EBI para ensinar comportamentos relacionados a delineamentos de
1593 sujeito único com aulas expositivas tradicionais (*lecture*). Os participantes foram
1594 aleatoriamente distribuídos entre os dois grupos que receberam as intervenções,

1595 e os resultados mostraram que os estudantes do grupo que recebeu a EBI
1596 conseguiram executar os comportamentos ensinados, mas não houve diferença
1597 significativa entre o desempenho deles e o desempenho dos alunos que
1598 assistiram às aulas tradicionais.

1599 Zinn, Newland e Ritchie (2015) também expandiram as áreas de
1600 aplicação do EBI ao realizar um estudo no qual 61 estudantes universitários
1601 foram ensinados s relações entre pares de propriedades e nomes de drogas
1602 genéricas usando procedimentos de ensino baseado em equivalência para o
1603 pareamento de estímulos auditivos e escritos. Foram utilizados dois grupos
1604 controle, nos quais a apresentação dos estímulos era aleatória, de maneira a
1605 não proporcionar a emergência de classes de equivalência. Um dos grupos
1606 controle foi baseado no critério da precisão, no qual os participantes deveriam
1607 atingir os mesmos critérios de proficiência do grupo experimental, possibilitando
1608 assim avaliar quantas tentativas necessitariam para conseguir o mesmo
1609 desempenho que o grupo que recebeu o ensino baseado em equivalência. O
1610 outro grupo controle foi baseado no número de tentativas, no qual os
1611 participantes poderiam fazer apenas a quantidade de tentativas utilizada pelo
1612 grupo experimental durante a intervenção, independente dos critérios de
1613 performance, permitindo avaliar qual desempenho atingiriam com a mesma
1614 quantidade de tentativas que o grupo que recebeu o ensino baseado em
1615 equivalências. Os resultados obtidos por Zinn, Newland e Ritchie (2015),
1616 mostram que os participantes do grupo controle por critério de precisão
1617 demonstraram requerer mais do que duas vezes mais tentativas para atingir os
1618 mesmos critérios de precisão do grupo experimental. Participantes do grupo
1619 controle por critério de tentativas obtiveram escores menores no pós-teste em
1620 todos os quatro tipos de medicamentos que estavam sendo ensinados, exceto
1621 pelos medicamentos do tipo estimulantes. A pesquisa de Zinn, Newland e Ritchie
1622 (2015) consiste em um avanço na área por mostrar que o uso de uma estratégia
1623 de ensino baseada na formação de classes de equivalência ensinou mais
1624 relações com menos quantidade de tentativas e maior precisão do que um
1625 método de ensino no qual as relações entre os estímulos eram ensinadas sem
1626 nenhuma ordem.

1627 Pytte e Fienup (2012) usaram EBI para ensinar repertórios relacionados
1628 a neuroanatomia em um contexto de sala de aula com 93 alunos. Os autores

1629 propuseram três questões a serem respondidas: (1) se os alunos seriam capazes
1630 de responder a relações simétricas sem que elas fossem diretamente treinadas;
1631 (2) se o uso de EBI produziria a emergência de respostas por meio da
1632 transitividade das relações; e (3) se o uso de EBI é tão eficiente quanto ensinar
1633 diretamente todas as relações entre classes de estímulos. Para as duas
1634 primeiras perguntas, o desempenho dos estudantes no pós-teste foi comparado
1635 com a probabilidade de acertarem as respostas do questionário ao acaso e foi
1636 observado que o desempenho dos estudantes era significativamente superior.
1637 Para responder à última questão, os autores compararam o desempenho de
1638 alunos em questões nas quais todas as relações transitivas entre estímulos
1639 haviam sido diretamente treinadas, com o desempenho em questões nas quais
1640 as relações transitivas não foram treinadas, e foi observado que o desempenho
1641 dos alunos nas duas formas não apresentou diferença estatisticamente
1642 significativa, ou seja, o uso de EBI foi tão efetivo quanto o treino direto de todas
1643 as relações. Com esses resultados, os autores concluíram que o uso de EBI foi
1644 bem-sucedido em contexto de uma sala de aula comum do ensino superior.

1645 Walker e Rehfeldt (2012) avaliaram o uso de EBI para ensinar
1646 delineamentos de sujeito único a estudantes do ensino superior via educação à
1647 distância. Foi utilizado um delineamento pré-teste/intervenção/pós-teste com
1648 avaliação de manutenção dos resultados em 16 semanas após o experimento,
1649 e, ainda que todos os participantes tenham sido bem-sucedidos em aprender as
1650 relações entre estímulos ensinadas, nem todos demonstraram manter essas
1651 relações 16 semanas depois do experimento, e, devido a isso, treinamento
1652 remediativo foi utilizado com sucesso para a maioria dos participantes.

1653 Critchfield (2014) utilizou EBI para ensinar estatística inferencial em um
1654 curso online. Contudo, como característica inovadora, o autor apresentou um
1655 procedimento para utilizar explicações escritas como forma de treinamento das
1656 relações entre estímulos ao invés de treinamento de comparação ao modelo
1657 (*matching to sample*). Utilizando um delineamento pré-teste/intervenção/pós-
1658 teste, no qual todos os 60 alunos participaram da intervenção, o autor observou
1659 que o desempenho dos alunos aumentou em 41%, na quantidade de respostas
1660 corretas e que eles completavam as lições que ensinavam as relações entre 28
1661 e 48 minutos, o que pode ser considerado um bom desempenho quando

1662 comparado com as duas horas que os alunos de Fienup e Crithcfield (2010)
1663 levaram para realizar o treino de comparação ao modelo.

1664 Albright, Reeve, Reeve e Kisamore (2015) avaliaram o uso de EBI para
1665 ensinar variabilidade estatística. Os autores também avaliaram se os
1666 comportamentos aprendidos por meio de procedimentos computadorizados
1667 seriam tão bem-sucedidos quanto a avaliações realizadas com papel e caneta.
1668 Por meio de um delineamento de pré-teste/intervenção/pós-teste, observou-se
1669 os alunos aprenderam as relações treinadas e emergentes, e que o desempenho
1670 dos alunos foi maior no pós-teste em ambos os tipos de avaliação
1671 (computadorizada e em papel e caneta) do que o desempenho pré-teste.

1672 Fienup, Wright e Fields (2015) conduziram dois experimentos para
1673 avaliar a participação de duas variáveis na eficiência do uso de EBI para ensinar
1674 neuroanatomia. No primeiro experimento, a efetividade de um protocolo de
1675 simples-ao-complexo (*simple-to-complex*) foi comparado com a efetividade de
1676 um protocolo de treinamento simultâneo, ao ensinarem classes de estímulos
1677 com 3 membros. No segundo experimento, as classes de estímulos possuíam
1678 quatro membros. Os resultados indicaram que o protocolo simples-ao-complexo
1679 foi mais eficiente para ensinar as classes de estímulos com 3 ou 4 membros do
1680 que o protocolo de treinamento simultâneo, indicando que a emergência
1681 imediata de relações entre classes de estímulos está em função interativa do tipo
1682 de protocolo de treinamento e do tamanho da classe de estímulos.

1683 O'Neill, Rehfeldt, Ninness, Muñoz e Mellor (2015) compararam a
1684 efetividade do uso de EBI, em um curso online, com um grupo de leitura dirigida
1685 (*assigned reading*) para ensinar a alunos de graduação a taxonomia skinneriana
1686 do comportamento verbal, por meio de um procedimento pré-
1687 teste/intervenção/pós-teste-manutenção. Contrariando os resultados anteriores,
1688 os alunos que participaram do grupo de EBI não tiveram um resultado superior
1689 do que o grupo de leitura em testes topográficos (nos quais o aluno deve
1690 escrever as respostas certas), mas houve aprimoramento no desempenho dos
1691 alunos que participaram do grupo de EBI quanto aos testes baseados em
1692 seleção da resposta correta.

1693 Reyes-Giordano e Fienup (2015) avaliaram o efeito do ensino de
1694 conceitos de neuratomia por meio de EBI no desempenho de alunos em testes
1695 baseados na topografia da resposta, por meio de um delineamento de *multiple*

1696 *probe baseline*. Os resultados mostraram que o desempenho dos estudantes
1697 aumentou tanto em testes baseados em topografia quanto em seleção de
1698 resposta. Além disso, os pesquisadores pediram a um professor de
1699 neuroanatomia, ingênuo quanto à pesquisa, que avaliasse as respostas dos
1700 alunos como em uma classe típica, e ele atribuiu notas que indicaram ganho no
1701 seu desempenho, o que sugere a efetividade do uso de EBI. Os resultados
1702 mostraram que os alunos no primeiro grupo controle precisaram de muito mais
1703 tentativas que do grupo de EBI para atingir o mesmo critério, enquanto que os
1704 alunos no segundo grupo controle, com a mesma quantidade de tentativas,
1705 tiveram uma proporção de acertos muito menor.

1706 Zinn, Newland e Ritchie (2015) avaliaram a efetividade e a eficácia do
1707 uso de EBI para ensinar a alunos relações entre propriedades e nomes de
1708 drogas genéricas. Dois grupos controle foram utilizados: o primeiro, deveria
1709 atingir o mesmo critério de proficiência do grupo EBI, mas com um número de
1710 tentativas livre; o segundo teria a quantidade de respostas corretas avaliadas,
1711 mas uma quantidade de tentativas limitada, pareada com o grupo EBI. Os alunos
1712 de ambos os grupos controle aprenderam relações entre estímulos
1713 apresentadas em uma ordem aleatória, em contraste com os alunos do grupo
1714 EBI, que receberam instrução organizada de maneira que relações emergentes
1715 ocorressem. Os alunos foram aleatoriamente distribuídos entre os grupos. Os
1716 resultados mostraram que os alunos que receberam EBI precisaram de menos
1717 tentativas para atingir os critérios de proficiência do que os alunos do grupo
1718 controle por critério, e tiveram um percentual de acertos maior do que os alunos
1719 pareados pelo grupo controle por tentativa.

1720 Albright, Schnell, Reeve e Sidener (2016) trouxeram mais para perto da
1721 própria análise do comportamento o uso da tecnologia de EBI ao utilizar uma
1722 estratégia baseada em equivalência, por meio de um software, para ensinar as
1723 relações entre quatro classes de estímulos representando funções do
1724 comportamento para dez alunos de graduação em um curso de análise do
1725 comportamento aplicada. Os participantes obtiveram melhora no desempenho
1726 médio de 89% na avaliação oral entre pré e pós intervenção, mantendo um
1727 desempenho de 87% na avaliação de manutenção realizada duas semanas
1728 depois da intervenção. Os participantes também apresentaram melhora na
1729 média de desempenho entre as avaliações pré e pós intervenção de 95% em

1730 testes escritos de múltipla escolha, e 95% na avaliação de manutenção, duas
1731 semanas após a intervenção.

1732 Fienup, Mylan, Brodksy e Pytte (2016) avaliaram o efeito de uma
1733 estratégia baseada em equivalência de estímulos para ensinar neuroanatomia
1734 funcional a alunos de uma disciplina de Neurociência Comportamental, e avaliar
1735 seu desempenho nas provas dessa disciplina. Os alunos que não participaram
1736 da intervenção obtiveram média de acertos nas questões da avaliação final da
1737 disciplina de 68,9% em questões não relacionadas àqueles tópicos utilizados no
1738 treino de equivalência, e 54,3% de respostas corretas em questões relacionadas
1739 ao treino de equivalência, enquanto que os participantes que passaram pelo
1740 treino obtiveram média de 75,6% de respostas corretas em questões não
1741 relacionadas ao treino de equivalência, e 77,6% de respostas corretas em
1742 questões relacionadas aos temas envolvidos no treino de equivalência.

1743 Greville, Dymond e Newton (2016) utilizaram EBI para ensinar
1744 neuroanatomia a estudantes, por meio de um delineamento pré-
1745 teste/intervenção/pós-teste, e avaliaram o desempenho dos alunos, bem como
1746 sua percepção de utilidade e confiança no tópico ensinado, em condições de
1747 complexidade crescente (o número de relações entre estímulos era aumentado
1748 em cada tópico ensinado). Os resultados mostraram que os alunos aprenderam
1749 as relações ensinadas, mostraram-se desejosos de mais materiais ensinados
1750 por meio de EBI, e que relataram se sentir mais confiantes quanto ao que
1751 aprenderam.

1752

1753

1754 **4.2 Avaliação quanto a elegibilidade a partir dos padrões WWC**

1755

1756

1757 Os artigos resultantes da busca e triagem foram avaliados quanto à sua
1758 elegibilidade para integrar os sumários de resultados promovidos pela WWC. Os
1759 artigos foram separados conforme o delineamento utilizado. Das 18 pesquisas,
1760 duas utilizaram delineamentos intrasujeito (*multiple probe design*), 8 utilizaram
1761 delineamentos intersujeitos, e 8 pesquisas utilizaram delineamentos que não se
1762 adequavam, exclusivamente, às características de inter ou intrasujeitos.

1763 As pesquisas que foram avaliadas como inelegíveis devido ao
 1764 delineamento assim o foram devido a não se adequarem aos critérios avaliativos
 1765 da WWC. As pesquisas a serem avaliadas por meio dos critérios para
 1766 delineamentos intersujeitos necessitam apresentar distribuição (aleatorizada ou
 1767 não) dos participantes entre grupos, e as pesquisas a serem avaliadas por meio
 1768 dos critérios para delineamentos intrasujeitos precisam, dentre outros critérios,
 1769 conter medidas repetidas e sequenciais da variável dependente (WWC, 2017b).
 1770 Os autores das pesquisas avaliadas como inelegíveis (Fienup et al., 2009;
 1771 Fienup & Critchfield, 2010; Walker, Rehfeldt, & Ninnes, 2010; Walker & Rehfeldt,
 1772 2012; Critchfield, 2014; Albright et al. 2015; Greville et al. 2016; e Albright et al.
 1773 2016), as classificaram como delineamentos pré-teste/intervenção/pós-teste,
 1774 mas não havia nelas distribuição dos participantes em grupos diferentes - e,
 1775 devido a isso, não foi possível avaliá-las como delineamento entre grupos.
 1776 Mesmo que se assemelhem a um delineamento intrasujeito, do tipo AB, a falta
 1777 de medidas do desempenho dos alunos em repetidas instâncias impossibilitou
 1778 que fossem avaliadas dessa forma, uma vez que apenas uma avaliação pré-
 1779 teste e uma pós-teste eram realizadas.

1780 Quanto às pesquisas avaliadas como delineamentos intersujeitos, a
 1781 Tabela 4 a seguir sumariza suas informações relevantes quanto à classificação
 1782 WWC. Nela, estão dispostos o autor e ano da pesquisa, com o tipo de
 1783 delineamento utilizado (*design*), os valores de atrito (geral/diferencial), o cálculo
 1784 de equivalência de linha de base [(ELB) quando necessário], o tamanho do efeito
 1785 da intervenção, e a significância estatística (*t*, bicaudal; $p=0,05$).

1786

1787

1788 **Tabela 4 Classificação das pesquisas com delineamento intersujeito**

Autor 1º	Ano	Design	Atrito	ELB	Tamanho de Efeito	Significância
Fields et al.	2009	RCT	4%/9%		$g = 2,2$	3,97
Fienup e Critchfield	2011	QED		$g = 0,01$	$g = 3,6$	5,8
Lovett et al.	2011	RCT	20%/40%	$g = 0,26$		
Pytte e Fienup	2012	QED				
Fienup et al.	2015	RCT				
O'Neil et al.	2015	QED		$g = 0,22$	$g = 0,73$	1,8
Zinn et al.	2015	RCT				
Fienup et al.	2016	QED		$g = 0,32$		

1789

1790 Das pesquisas avaliadas por meio dos padrões de elegibilidade WWC
1791 para delineamento intersujeito, apenas uma foi classificada como “Elegível para
1792 atender aos padrões WWC para Delineamentos de Grupo Sem Reservas”
1793 (destacada em verde, na Tabela 4): Fields et al. (2009). Duas pesquisas foram
1794 avaliadas como Elegíveis para Atender aos Padrões WWC para Delineamentos
1795 de Grupo Com Reservas (destacadas em roxo, na Tabela 4): Fienup e Critchfield
1796 (2011), e O’Neil et al. (2015).

1797 Das pesquisas classificadas como elegíveis, apenas Fields et al. (2009)
1798 é um caso de RCT, considerado padrão ouro da área. A pesquisa apresentou
1799 um atrito considerado baixo, e, portanto, não precisou passar por avaliação de
1800 equivalência de linha de base. Fienup e Critchfield (2011) foi o caso de pesquisa
1801 QED com valor de equivalência de linha de base aceito dentro dos critérios WWC
1802 (inferior a 0,05). O’Neil et al. (2015), também um caso de pesquisa QED,
1803 apresentou um valor de equivalência de linha de base próximo do valor de
1804 exclusão (acima de 0,25), mas pode ser classificada como “Elegível”.

1805 As pesquisas restantes foram classificadas como “Não atendem aos
1806 padrões da WWC para Delineamentos de Grupo”, em função de fatores diversos.
1807 Lovett et al., (2011) consistiu em um caso de RCT, mas com um valor de atrito
1808 que ameaçou a validade interna da pesquisa. Contudo, uma vez que os autores
1809 não informaram o valor exato de alunos que abandonaram o grupo controle,
1810 estimamos, para esta revisão, um valor que equiparasse o grupo controle com o
1811 grupo experimental, de maneira que fosse possível estimar a quantidade de
1812 desertores, e, assim, calcular o atrito - que, neste caso, foi alto. A partir disso, foi
1813 calculado o valor de equivalência de linha de base, que ultrapassou os limites
1814 permitidos para classificar a pesquisa como “Elegível Com Reservas” (0,25).

1815 Pyte e Fienup (2012) trata-se de uma pesquisa com três experimentos
1816 distintos. Enquanto nos dois primeiros não houve distribuição dos participantes
1817 em grupos distintos, no terceiro experimento o repertório dos alunos foi avaliado
1818 de maneira que era possível separar os comportamentos que foram treinados
1819 via EBI e aqueles que não foram, mas como os autores não forneceram
1820 informações suficientes para se verificar a equivalência de linha de base (e,
1821 depois, o tamanho do efeito), a pesquisa foi classificada como inelegível.

1822 A pesquisa conduzida por Fienup et al. (2015) foi classificada com RCT,
1823 mas foi avaliada como inelegível devido ao fato de que, apesar dos autores

1824 distribuírem os participantes aleatoriamente entre os grupos, ambos os grupos
1825 receberam a intervenção, ainda que com propriedades diferentes. Dessa forma,
1826 não houve um grupo controle com o qual o desempenho dos participantes do
1827 grupo experimental fosse comparado, e isso torna a pesquisa uma comparação
1828 da efetividade de diferentes parâmetros da intervenção (EBI). Por isso, mesmo
1829 que o desempenho dos participantes fosse comparado, essa comparação não
1830 seria representativa do contraste do desempenho dos alunos que passaram pela
1831 intervenção, uma vez que ambos os grupos passaram por ela.

1832 A pesquisa de Zinn et al. (2015), foi classificada como inelegível porque
1833 os autores não apresentaram os dados dos resultados da pesquisa de maneira
1834 que fosse possível avaliá-los nesta revisão. É fato que os dados do desempenho
1835 dos grupos foram apresentados graficamente, mas obtê-los a partir dos gráficos
1836 geraria valores pouco confiáveis em relação à precisão.

1837 Por fim, a pesquisa de Fienup et al. (2016), foi avaliada como inelegível
1838 devido ao valor de equivalência de linha de base entre os grupos (de 0,32),
1839 considerado acima do limite aceito pela WWC.

1840 Quanto às pesquisas avaliadas como delineamento intrasujeitos, ambas
1841 utilizaram delinamentos de *multiple probe designs*. Fienup, Covey e Critchfield,
1842 (2010) foi avaliada como inelegível devido ao fato dos pesquisadores não
1843 apresentarem informações quanto à concordância interobservadores, requisito
1844 preliminar para esse tipo de delineamento. Tal problema não ocorre no trabalho
1845 de Reyes-Giordano e Fienup (2010), mas, apesar de apresentarem valores de
1846 concordância interobservadores, a pesquisa foi classificada como inelegível por
1847 não apresentar pontos de dados suficientes para atender os padrões WWC, e
1848 nem mesmo gráficos de frequência acumulada que permitissem a análise visual
1849 dos resultados de acordo com os critérios propostos para avaliar pesquisas
1850 intrasujeito. Ao invés disso, os autores (Reyes-Giordano & Fienup, 2010)
1851 apresentaram as informações dos resultados como tradicionalmente se faz em
1852 pesquisas intersujeitos, comparando o desempenho dos alunos antes e depois
1853 intervenção.

1854

1855

1856 **4.3 Síntese quantitativa dos resultados**

1857

1858 4.3.1 *Tamanho de efeito*

1859

1860 A WWC (2017a) recomenda o cálculo do tamanho de efeito das
1861 intervenções como forma de comparar adequadamente o impacto de uma
1862 intervenção em diferentes grupos, realizada em pesquisas diferentes. Das
1863 pesquisas avaliadas como Elegíveis para Atender aos Padrões WWC de
1864 Delineamentos de Grupo (com ou sem reservas), o tamanho de efeito foi
1865 calculado, como é possível observar na Tabela 4. Há de se notar, no entanto,
1866 que Fields et al. (2009) não apresentou as informações necessárias para o
1867 cálculo do tamanho de efeito por parte desta revisão, mas apresentou um valor
1868 de tamanho de efeito calculado pelos próprios autores - que geralmente é aceito,
1869 segundo recomendação da WWC (2017a). Esse valor necessitou ser alterado,
1870 tendo em vista que o cálculo recomendado pela WWC (2017a) é uma variação
1871 do padrão (d , de Cohen) por conter um ajuste devido a pequenas amostras, por
1872 meio da fórmula de g de Hedges. Isto posto, o valor apresentado por Fields et
1873 al. (2009), era de $d = 2,3$, e foi ajustado para $g = 2,2$.

1874 Outra observação necessária diz respeito à diferença quanto aos valores
1875 de tamanho de efeito obtidos a partir dos resultados apresentados por O'Neil et
1876 al. (2015) e Fields et al. (2009) junto ao valor apresentado por Fienup e Critchfield
1877 (2011). A grande diferença entre esses valores pode ser devida ao fato de que,
1878 em Fields et al. (2009) e Fienup e Critchfield (2011), o tamanho de efeito foi
1879 calculado comparando-se o desempenho do grupo de intervenção com um grupo
1880 controle no qual os participantes não realizaram nenhuma atividade até o pós-
1881 teste, enquanto O'Neil et al. (2015) comparou o desempenho dos estudantes
1882 que receberam EBI com o desempenho de estudantes que realizaram atividades
1883 de leitura dirigida.

1884 Considerando a disparidade de condições da pesquisa de O'Neil et al.
1885 (2015), é possível apresentar duas médias diferentes de tamanho de efeito para
1886 o uso de EBI como estratégia didática no ensino superior. O que poderíamos
1887 chamar de um tamanho de efeito puro (obtido quando a intervenção é
1888 comparada com o desempenho de participantes que não realizaram nenhuma
1889 atividade), calculado pela média do tamanho de efeito das pesquisas de Fields
1890 et al. (2009) e Fienup e Critchfield (2011), no valor de $g = 2,9$; e um tamanho de
1891 efeito geral, calculado a partir da média do tamanho de efeito de todas as

1892 pesquisas elegíveis, no valor aproximado de $g = 2,18$. Em ambos os casos, o
1893 valor pode ser considerado alto (acima de 0,8).

1894

1895 4.3.2 *Índice de aprimoramento*

1896

1897 A distinção entre três possíveis tamanhos de efeito apresentada
1898 anteriormente nos permite calcular três possíveis índices de aprimoramento a
1899 partir de três cenários possíveis: (A) o tamanho de efeito geral, calculado a partir
1900 da média das três pesquisas avaliadas como Elegíveis; (B) o tamanho de efeito
1901 puro, calculado a partir da média das duas pesquisas que utilizaram grupo
1902 controle sem nenhuma atividade; e (C) o tamanho de efeito da EBI comparada
1903 com leitura dirigida.

1904 Considerando essas três possibilidades, o índice de aprimoramento do
1905 cenário A, é de 48%, indicando que se espera que um aluno de desempenho
1906 médio, ao participar de uma classe cuja didática seja EBI, melhore seu
1907 desempenho em 48%. Por exemplo, pode-se esperar que um aluno cujo
1908 desempenho em uma disciplina tenha o colocado em reprovação (com uma nota
1909 de 40% na média geral, pelos padrões de algumas instituições de ensino
1910 superior), poderia atingir um desempenho de 88%, caso recebesse EBI como
1911 forma de instrução; o suficiente para que fosse aprovado (com uma nota acima
1912 de 70%). Esse valor do índice de aprimoramento também indica que, por meio
1913 do uso de EBI, pode-se esperar que 98% dos alunos teriam um desempenho
1914 acima da média do desempenho de alunos que não participassem de aulas com
1915 EBI.

1916 No cenário B, o índice de aprimoramento é apenas um ponto percentil
1917 maior do que o cenário A, e a mesma análise pode ser aplicada. Contudo, como
1918 será discutido mais com mais detalhes posteriormente, como as pesquisas que
1919 foram utilizadas para calcular o tamanho de efeito do cenário B (e o índice de
1920 aprimoramento derivado) ensinaram apenas repertórios relacionados à
1921 estatística, seria mais adequado afirmar que o índice de aprimoramento
1922 respectivo (49%) é aplicável ao ensino de estatística - ou seja, os ganhos no
1923 desempenho indicados por esse índice de aprimoramento devem ser esperados
1924 quando se compara o desempenho de alunos que passaram por um curso via

1925 EBI para aprender repertórios relacionados à estatística com o desempenho de
1926 alunos que não passaram por nenhum curso de estatística.

1927 No cenário C, o índice de aprimoramento é de 27%, indicando que um
1928 aluno de desempenho médio que participe de um curso cuja estratégia de ensino
1929 seja EBI apresentará um desempenho 27% superior a um aluno que participou
1930 o mesmo curso, mas cuja estratégia de ensino era a leitura dirigida. Além disso,
1931 esse índice de aprimoramento aponta que 77% dos alunos de uma turma de EBI
1932 terão desempenho maior do que a média do desempenho de alunos de uma
1933 turma de leitura dirigida.

1934

1935 4.3.3 *Significância estatística*

1936

1937 Os valores de significância estatística foram calculados de acordo com
1938 os parâmetros apresentados pela WWC (2017a). Para Fields et al. (2009), a
1939 diferença observada no desempenho entre o grupo experimental e o grupo
1940 controle é estatisticamente significativa ($t = 3,97$; $df = 19$; $p = 0,05$). A diferença
1941 no desempenho entre o grupo experimental e o grupo controle em Fienup e
1942 Crithcfield (2011), também se mostrou estatisticamente significativa ($t = 5,8$; $df =$
1943 26 ; $p = 0,05$). Entretanto, a diferença no desempenho entre o grupo experimental
1944 e o grupo controle em O'Neil et al. (2015) não foi estatisticamente significativa (t
1945 $= 1,8$; $df = 24$; $p = 0,05$).

1946

1947

1948 4.4 **Síntese qualitativa dos resultados**

1949

1950 Quanto à caracterização dos resultados das pesquisas avaliadas como
1951 Elegíveis, a WWC as classifica como:

1952

- 1953 • Fields et al. (2009): Efeito positivo estatisticamente significativo.
- 1954 • Fienup e Crithcfield (2011): Efeito positivo estatisticamente significativo.
- 1955 • O'Neil et al. (2015): Efeito positivo substancialmente importante.

1956

1957 A partir dos critérios utilizados para determinar a classificação da
1958 efetividade de uma intervenção por meio de todos os estudos avaliados como

1959 “Elegíveis”, os resultados do uso de EBI como estratégia didática no ensino
1960 superior podem ser caracterizados tanto como Efeitos positivos (forte evidência
1961 de um efeito positivo com nenhuma evidência contrária que a supere) e Efeitos
1962 potencialmente positivos (evidência de um efeito positivo com nenhuma
1963 evidência contrária que a supere). Contudo, a recomendação é de que, caso
1964 critérios múltiplos sejam atingidos, a classificação mais alta seja adotada.

1965 Em relação à extensão das evidências, os resultados das pesquisas
1966 revistas neste estudo fornecem evidências para uma extensão pequena dos
1967 resultados, uma vez que atendem os dois primeiros critérios da Tabela 3, mas
1968 não atendem o último - isto é, no mínimo 350 estudantes participando do total de
1969 amostras, já que, somando os participantes das três pesquisas, temos um total
1970 de 75 participantes.

1971 Por fim, para a síntese que compõe o guia prático, as seguintes
1972 informações são relevantes:

- 1973 • O número de estudos elegíveis foi três.
- 1974 • Apenas um desses estudos foi avaliado como “Elegível” para
1975 Atender aos Padrões WWC para Delineamentos de Grupo Sem
1976 Reservas. Os outros dois estudos Atendem aos Padrões, mas
1977 Com Reservas.
- 1978 • Os estudos são representativos da abrangência de participantes,
1979 contextos e comparações para a recomendação da adoção da
1980 EBI como estratégia de ensino, ainda que a extensão da
1981 evidência seja pequena.
- 1982 • Os resultados dos estudos podem ser atribuídos à prática
1983 recomendada.
- 1984 • Os resultados dos estudos são consistentemente positivos.
- 1985

1986 5 DISCUSSÃO

1987

1988

1989 Três resultados principais são destacados para a discussão neste
1990 capítulo. O primeiro deles diz respeito à recomendação do uso de EBI como
1991 estratégia didática no ensino superior como uma prática baseada em evidências,
1992 principalmente quanto às implicações e limitações dessa recomendação. No
1993 segundo tópico, levamos em consideração os artigos excluídos dos resultados
1994 considerados pela WWC para discutir sobre as consequências da adoção dos
1995 critérios dessa instituição para a avaliação de PBE. No terceiro tópico, usamos
1996 os resultados desta pesquisa para refletir sobre algumas implicações quanto às
1997 práticas culturais tradicionais da comunidade de analistas do comportamento em
1998 relação à pesquisa científica.

1999

2000

2001 **5.1 Recomendação e limitações da EBI como uma prática baseada em** 2002 **evidências**

2003

2004

2005 A partir dos dados obtidos por meio da avaliação das pesquisas quanto
2006 ao uso de EBI como estratégia didática no ensino superior, é possível concluir
2007 que, de acordo com os critérios da WWC para avaliação de intervenções na área
2008 da educação, a EBI pode ser recomendada como uma prática baseada em
2009 evidências, apresentando fortes evidências de efeitos positivos com nenhuma
2010 evidência contrária que as superem, embora apresentando uma extensão
2011 pequena das evidências obtidas.

2012

2013 Os resultados desta revisão mostram que, de acordo com os padrões da
2014 WWC, o uso de EBI tem um grau elevado de efetividade para o ensino de
2015 comportamentos acadêmicos a alunos do ensino superior. O tamanho de efeito
2016 geral obtido, de $g = 2,18$, ultrapassa em mais do que duas vezes o critério
2017 convencional, de $g = 0,8$, para se considerar um tamanho de efeito alto. Como
2018 apresentado anteriormente, esse valor é o suficiente para estimar que um aluno
2019 de desempenho médio que passe por um curso via EBI atinja um desempenho
48% maior do que outro aluno que não participasse desse curso, e que, no caso

2020 de um aluno com notas baixas, o curso poderia melhorar seu desempenho o
2021 suficiente para que fosse aprovado.

2022 A partir desses dados, poderíamos recomendar o uso de EBI como
2023 estratégia didática principal de uma disciplina de graduação, substituindo as
2024 tradicionais aulas expositivas ou leituras dirigidas, e, com isso, garantir que
2025 praticamente todos os alunos aprendessem os comportamentos ensinados.
2026 Trata-se de um cenário idealizado, mas louvável, na medida em que os dados
2027 obtidos nesta revisão consistem em uma evidência confiável da aplicação dessa
2028 estratégia didática, que supera a falta de informações quanto à eficácia de outras
2029 estratégias didáticas utilizadas no ensino superior. Entre dar uma aula
2030 expositiva, sem saber qual a efetividade desse método, e dar uma aula via EBI,
2031 sabendo de sua eficácia, esta segunda opção parece preferível.

2032 Uma alternativa, também recomendável, seria o uso de EBI como
2033 estratégia didática auxiliar. Ela poderia ser utilizada, por exemplo, como forma
2034 de melhorar o desempenho de alunos abaixo da média, ou à beira da
2035 reprovação. Como os resultados mostraram, o efeito da implementação de EBI
2036 é alto o suficiente para melhorar o desempenho de um aluno que iria reprovar,
2037 de maneira que ele atinja os critérios de aprovação adotados em muitas
2038 universidades brasileiras. Isso pode consistir em uma decisão pedagógica mais
2039 interessante do que pedir a um aluno que estude novamente para os exames
2040 finais, utilizando os mesmos métodos de ensino por meio dos quais não
2041 conseguiu desempenho suficiente para ser aprovado em primeira instância.

2042 Como estratégia didática auxiliar, professores poderiam pedir aos alunos
2043 que realizassem atividades programadas via EBI em seus computadores
2044 pessoais (ou mesmo em aplicativos de celular), em contextos fora da sala de
2045 aula. E, como é possível estimar que quase todos os alunos de uma turma de
2046 EBI atinjam um desempenho acima da média, quando comparados com alunos
2047 que não participassem das aulas com EBI, seria de esperar que eles viessem à
2048 aula com os comportamentos já aprendidos. Isso liberaria tempo do professor,
2049 que, ao invés de ensinar repertórios básicos, poderia se dedicar a
2050 comportamentos mais complexos, como propôs Skinner (1968), por meio de
2051 outros métodos didáticos.

2052 Isso não significa que o uso de EBI deva ser limitado a comportamentos
2053 simples ou básicos. Repertórios estatísticos foram os mais frequentemente

2054 ensinados via EBI, de acordo com as pesquisas encontradas nesta revisão. O
2055 segundo repertório mais frequente foi neuroanatomia. Esses dois repertórios
2056 foram escolhidos por serem considerados bastante complexos, e por muitos
2057 alunos os considerarem difíceis de aprender (Field et al., 2009; Fineup et al.
2058 2010; Fienup & Critchfield, 2011). Considerando-se que a maioria das disciplinas
2059 de análise do comportamento, nas graduações em psicologia brasileiras,
2060 ensinam os princípios básicos do comportamento, as evidências apresentadas
2061 nesta pesquisa sugerem que o uso de EBI foi efetivo para ensinar
2062 comportamentos ainda mais complexos do que princípios básicos, resultantes
2063 da aprendizagem da taxonomia dos operantes verbais clássicos, como
2064 mostraram O'Neil et al. (2015).

2065 A efetividade demonstrada quanto ao uso de EBI como estratégia para
2066 o ensino de comportamentos que podem ser considerados complexos responde
2067 parte das críticas feitas por educadores, conforme apresentadas por Bernstein e
2068 Chase (2013), de que as estratégias didáticas analítico-comportamentais
2069 ensinam apenas comportamentos simples. Se juntarmos a isso o fato de que os
2070 comportamentos produzidos por procedimentos de formação de classes de
2071 equivalência entre estímulos têm sido associados ao que outras áreas
2072 chamariam de habilidades cognitivas superiores (De Rose, 1993; Matos, 1999;
2073 Rabelo & De Rose, 2015), então podemos considerar que a EBI está em um
2074 posição muito favorável para ser difundida, implementada na educação superior,
2075 e recomendada como uma PBE.

2076 Não se sabe, no entanto, os limites da complexidade dos
2077 comportamentos que podem ser ensinados via EBI. Como será discutido mais
2078 detidamente em breve, a extensão das evidências encontradas quanto ao uso
2079 de EBI não levou a respostas para questões acerca dos limites das propriedades
2080 dos comportamentos acadêmicos que podem ser ensinados usando esses
2081 procedimentos. Mais pesquisas são necessárias para responder questões,
2082 como, por exemplo, se EBI seria uma estratégia efetiva para ensinar cadeias
2083 extensas de comportamento, como o planejamento completo, condução e
2084 relatório de uma pesquisa científica; se repertórios de matemática avançada, ou
2085 de filosofia, em nível de doutorado, poderiam ser ensinados efetivamente, por
2086 meio de EBI; se a EBI ainda manteria sua efetividade caso um curso inteiro fosse
2087 implementado utilizando-a como didática principal.

2088 O entusiasmo com que poderíamos recomendar a implementação de
2089 cursos no ensino superior que utilizassem EBI, deve, no entanto, ser moderado
2090 e relativizado, devido a outros resultados apresentados nesta revisão. O uso de
2091 EBI como estratégia didática no ensino superior foi avaliada como tendo Efeitos
2092 Positivos, quanto à sua classificação de efetividade, mas com extensão pequena
2093 de evidências. As pesquisas que atingiram os critérios avaliativos apresentaram
2094 apenas 21% da quantidade de alunos necessária para se atingir o critério de
2095 extensão média ou ampla.

2096 Nesse sentido, os resultados desta pesquisa corroboram, no entanto, os
2097 dados de Stockard e Wood (2016), de que a maioria dos sumários de evidências
2098 da WWC é baseada em apenas um estudo com boa avaliação, e envolve dados
2099 de poucos alunos. Esse problema se deve, de acordo com os autores, ao fato
2100 de que os critérios de seleção e avaliação da WWC utilizam a lógica de limites
2101 que devem ser alcançados pelo estudo, e que excluem automaticamente
2102 aqueles que não atingem tais limites. Devido a isso, a maior parte dos estudos
2103 disponíveis na literatura científica sobre uma determinada intervenção acaba
2104 sendo descartada por não passar em um ou outro critério, reduzindo
2105 drasticamente o número de estudos considerados de qualidade à disposição.
2106 Voltaremos a esse problema adiante, ao discutirmos as pesquisas que foram
2107 excluídas dos resultados desta revisão.

2108 Tão importante quanto a quantidade total de alunos que participaram das
2109 pesquisas talvez seja a diversidade de contextos em que a intervenção é
2110 aplicada. Os critérios para avaliação da extensão dos dados da intervenção,
2111 propostos pela WWC (2017a), requerem, no mínimo, dois estudos e dois
2112 cenários, além do número total de 350 estudantes utilizados como amostra nas
2113 pesquisas. A avaliação do uso de EBI, resultante desta revisão, resultou em
2114 exatamente três estudos e dois contextos de aplicação (*settings*), o ensino de
2115 estatística e taxonomia do comportamento verbal. Considerando-se que os
2116 resultados sobre o ensino de taxonomia de operantes verbais foi prejudicado
2117 devido ao resultado falho quanto à significância estatística, os dois estudos
2118 restantes apresentam evidências quanto ao ensino de estatística, por meio de
2119 EBI, a alunos do ensino superior.

2120 Assim, afirmar que o uso de EBI é efetivo para ensinar repertórios
2121 acadêmicos a alunos do ensino superior é uma extrapolação arriscada, mas

2122 permitida segundo os critérios apresentados pela WWC (2017a, 2017b). A
2123 justificativa inerente aos critérios da instituição é que, se estudos suficientes
2124 apresentam validade interna suficiente para atender aos altos padrões exigidos,
2125 então os resultados apresentam confiabilidade suficiente para serem adotados
2126 em diversos contextos. Essa lógica, no entanto, é discutível. É necessário que a
2127 efetividade de uma intervenção seja avaliada, com a devida qualidade
2128 metodológica, em contextos diversos para que se possa afirmar que ela é efetiva
2129 em contextos diversos, e, portanto, pode ser adotada amplamente.

2130 A WWC (2017a) fornece parâmetros cumulativos para avaliar a
2131 abrangência da efetividade de uma intervenção, por meio dos critérios das
2132 Tabelas 1, 2 e 3 (apresentadas no capítulo Método), de maneira que, quanto
2133 mais pesquisas que atendam aos critérios avaliativos sejam encontradas, melhor
2134 se torna a recomendação daquela intervenção⁷. Essa característica torna mais
2135 flexível e dinâmica a recomendação de uma prática, permitindo que ela seja
2136 melhorada ou piorada conforme mais pesquisas são identificadas, avaliadas e
2137 integradas aos resultados. O problema que destacamos, no entanto, não é esse,
2138 mas o fato de que, com os critérios vigentes, uma prática como o uso de EBI
2139 para o ensino de comportamentos a alunos do ensino superior pode ser
2140 recomendada com mais abrangência de aplicação (todo o ensino superior) do
2141 que as evidências obtidas apoiam (ensino de estatística em disciplinas de
2142 graduação).

2143 Problemas como esse podem ser atribuídos ao enfoque dado pela WWC
2144 à avaliação da literatura em termos exclusivos de validade interna das pesquisas.
2145 Uma crítica semelhante a essa é apresentada por Stockard e Wood (2016), ao
2146 afirmar que a utilização de dados agregados, como a média do tamanho de efeito
2147 das intervenções, para a elaboração de sumários de evidências, pode acabar
2148 mascarando variações importantes entre elementos amostrais ou da análise dos
2149 estudos. No caso dos resultados deste estudo, o uso de dados agregados
2150 reporta a efetividade de uma estratégia de ensino que deveria ser limitada a um
2151 contexto, mas que, por ter atingido aos critérios avaliativos da WWC, pode ser
2152 recomendada com uma abrangência maior do que as pesquisas de fato indicam.

⁷ É possível encontrar no site da instituição muitos exemplos de classificação da efetividade (no coluna, *effectiveness rating*) de uma intervenção, e constatar como essa classificação muda conforme a quantidade de pesquisas encontradas que atendem aos padrões avaliativos.

2153 Quanto a esse tipo de problema, Stockard e Wood (2016) sugerem que
2154 uma análise mais tradicional, como os sumários de resultados empregados pelas
2155 ciências sociais, poderia oferecer informações mais abrangentes e precisas
2156 quanto aos efeitos da aplicação de uma intervenção, como a extensão das
2157 variações de resultados, a persistência dos efeitos da intervenção com o passar
2158 o tempo, e outras características relacionadas à validade externa, como
2159 características dos estudantes, escolas, comunidades. Por meio de sumários
2160 mais abrangentes, como na proposta de Stockard e Wood (2016), os resultados
2161 da presente pesquisa poderiam ser comunicados à comunidade geral de
2162 maneira mais precisa quanto aos limites da extensão de sua aplicabilidade,
2163 corrigindo o erro de generalização que destacamos. Além disso, seria possível
2164 relatar muito mais informações, que em nosso estudo foram obtidas mas
2165 descartadas, como discutido a seguir.

2166

2167

2168 **5.2 Sobre os limites dos padrões avaliativos adotados pelas PBE**

2169

2170

2171 Os limites que compõem os critérios avaliativos e padrões da WWC
2172 restringem a quantidade de informações que constam nas revisões sistemáticas
2173 de literatura. Nesse sentido, os resultados desta pesquisa corroboram o
2174 argumento apresentado por Stockard e Wood (2016), de que o modelo de
2175 triagem e avaliação de literatura por limites excludentes acaba reduzindo a
2176 quantidade de pesquisas a níveis que não são representativos da literatura
2177 disponível. Segundo os autores, esse problema pode enviesar a informação
2178 proporcionada pela WWC quanto ao efeito de uma prática. De acordo com a
2179 pesquisa conduzida pelos autores (Stockard e Wood, 2016), que avaliou a
2180 aplicação dos critérios WWC para revisar procedimentos bem estabelecidos e
2181 de efetividade demonstrada na literatura educacional sobre alfabetização,
2182 menos de 40% da literatura disponível passou pelos critérios.

2183 Em nosso estudo, das 18 pesquisas que passaram nos critérios de
2184 triagem da literatura sobre o uso de EBI no ensino superior, oito pesquisas foram
2185 excluídas devido ao delineamento utilizado, inadequado para avaliação pelos
2186 procedimentos da WWC por não se caracterizarem nem como inter, nem como

2187 intrasujeitos, e sequer tiveram seus resultados considerados (Fienup et al., 2009;
2188 Fienup & Critchfield, 2010; Walker, Rehfeldt & Ninnes, 2010; Walker & Rehfeldt,
2189 2012; Critchfield, 2014; Albright et al. 2015; Greville et al. 2016; e Albright et al.
2190 2016). Além dessas, outras sete pesquisas (Fineup et al., 2010; Lovett et al.,
2191 2011; Pytte e Fienup, 2012; Reyes-Giordano e Fienup, 2015; Zinn et al., 2015;
2192 Fienup et al., 2015; e Fienup et al., 2016) tiveram seus resultados excluídos
2193 devido a problemas de validade interna, identificados a partir do uso dos critérios
2194 avaliativos da WWC (2017b). Dessa forma, apenas 17% dos artigos que
2195 passaram pelos critérios de triagem estabelecidos no protocolo desta revisão
2196 tiveram seus resultados computados nas sínteses de efeito da intervenção
2197 pesquisada.

2198 Por se fundamentarem em uma área já bem estabelecida, com boas
2199 evidências acerca da estabilidade das relações funcionais entre procedimentos
2200 de formação de classes de equivalência e emergência de comportamentos
2201 treinados diretamente ou não, muitos dos resultados das pesquisas excluídas ou
2202 avaliadas como “Não Atendem aos Padrões WWC” poderiam ser aproveitadas
2203 de alguma forma. Algumas dessas pesquisas expandem as áreas de aplicação
2204 da EBI ao testar seu uso para o desenvolvimento de outros repertórios, além de
2205 estatística (Fineup, Covey, & Critchfield, 2010; Greville, Dymond, & Newton,
2206 2016; Pytte, & Fienup, 2012; Reyes-Giordano, & Fienup, 2015; Walker, Rehfeldt
2207 & Ninnes, 2010; Lovett et al., 2011; Walker & Rehfeldt, 2012; Albright et al.,
2208 2016). Outras avaliam a adequabilidade da transposição do uso de EBI para
2209 contextos mais semelhantes à sala de aula (Fienup, & Critchfield, 2011; Fienup
2210 et al., 2016). Outros ainda avaliam parâmetros e componentes do procedimento
2211 de EBI (Fineup, Wright, & Fields, 2015; Zinn et al., 2015).

2212 A partir da proposta, defendida anteriormente no Capítulo 1, de que, sob
2213 a perspectiva do behaviorismo radical, a melhor evidência consiste na replicação
2214 das pesquisas, então é possível considerar que o resultado das pesquisas
2215 excluídas desta revisão poderia agregar informações importantes, na medida em
2216 que elas replicam, sob parâmetros diferentes, a aplicação de EBI no ensino
2217 superior, acrescentando, portanto, validade à efetividade dessa intervenção.
2218 Vale notar também que todas as pesquisas que compuseram o resultado da
2219 triagem da literatura relataram efeitos positivos da intervenção no desempenho
2220 dos alunos.

2221 Seria um desperdício simplesmente descartar o conhecimento
2222 acrescentado por essas pesquisas à área de equivalência de estímulos, como
2223 um todo, e à área de EBI, mais especificamente. Mesmo que elas apresentem
2224 problemas de validade interna, isso não significa que as informações
2225 apresentadas em seus resultados sejam irrelevantes. Esses resultados
2226 poderiam ser tratados como possibilidades abertas, para as quais alguma
2227 informação já foi obtida, mas para as quais mais pesquisas (replicações) são
2228 necessárias.

2229 Quanto ao modelo de agregação de informações, a tradição analítico-
2230 comportamental não é compatível com aquele promovido pela WWC e adotado
2231 pelas PBE, no sentido de que, considerando a replicação como melhor forma de
2232 evidência, a agregação dos resultados das pesquisas, ao invés de sua exclusão,
2233 é fundamental para o fortalecimento das evidências de uma intervenção. Isso
2234 não implica, é claro, afirmar que a tradição analítico-comportamental admite o
2235 acúmulo indiscriminado de informações, independentemente da qualidade
2236 metodológica das pesquisas, mas que seus parâmetros de inclusão de
2237 resultados são diferentes e próprios dessa ciência.

2238

2239

2240 **5.3 Sobre o uso de delineamentos intersujeitos por analistas do** 2241 **comportamento**

2242

2243

2244 No Capítulo 1 foi discutido sobre o problema gerado pelo uso indistinto
2245 dos dois sentidos da expressão “práticas baseadas em evidências”. Ao mesmo
2246 tempo em que se questiona se existem evidências bem fundamentadas acerca
2247 da relação entre a implementação de uma intervenção e os resultados
2248 produzidos (sentido epistemológico), também se questiona acerca de evidências
2249 bem fundamentadas que justifiquem em uma análise de custo-benefício, a
2250 recomendação da adoção geral de uma prática, ou seu estabelecimento como
2251 política pública (sentido atuarial).

2252 Diante desse contraste, foi destacado que analistas do comportamento,
2253 de acordo com os compromissos filosóficos adotados pelo Behaviorismo
2254 Radical, têm, tradicionalmente, utilizado delineamentos de sujeito único como

2255 estratégia metodológica para adquirir evidências sobre relações funcionais entre
2256 eventos do ambiente e o comportamento (humano ou não-humano), e adotado
2257 a replicação como critério de “melhor evidência”. Dessa forma, no sentido
2258 epistemológico, haveria boas evidências acerca de algumas das intervenções
2259 produzidas por analistas do comportamento, uma vez que muitas delas possuem
2260 uma considerável história de replicações e derivam de princípios
2261 comportamentais bem consolidados.

2262 Contudo, questionamos se, em um sentido atuarial, a comunidade
2263 analítico-comportamental dispõe de evidências que apoiem a magnitude
2264 esperada das mudanças proporcionadas pela implementação de intervenções
2265 propostas por analistas do comportamento - uma vez que, como foi discutido,
2266 estes são críticos quanto à validade do uso de delineamentos de pesquisa com
2267 grupos como forma de descrever relações funcionais.

2268 Os resultados apresentados nesta revisão podem ser tomados como um
2269 exemplo de que analistas do comportamento têm conduzido pesquisas bem-
2270 sucedidas com delineamentos de grupo. Estudos como os de Pytte e Fienup
2271 (2012) e Fienup et al. (2016) utilizaram delineamentos de grupo para avaliar a
2272 adequabilidade da implementação do ensino por meio de EBI em ambientes
2273 próximos da sala de aula. Pesquisas como as de Lovett et al. (2011), O’Neill et
2274 al. (2015) e Zinn et al. (2015) também utilizaram esse tipo de delineamento para
2275 comparar a efetividade do ensino por meio de EBI com a efetividade de outras
2276 estratégias didáticas. Mesmo que, pelos critérios da WWC, uma proporção
2277 diminuta dessas pesquisas tenha sido avaliada positivamente, uma quantidade
2278 suficiente foi avaliada com tendo qualidade para a recomendação da prática
2279 como baseada em evidências.

2280 A avaliação positiva recebida por uma estratégia de ensino analítico-
2281 comportamental a partir de critérios aplicados a pesquisas de grupo pode ser
2282 vista como um incentivo para que analistas do comportamento utilizem, com
2283 mais frequência, delineamentos intersujeitos. Esse tipo de incentivo à variação
2284 de práticas culturais por parte de analistas do comportamento não deve ser visto,
2285 no entanto, como uma ameaça às práticas de pesquisa tradicionalmente
2286 adotadas, mas como uma forma de expandir a fonte (comportamento de
2287 indivíduos em grupos) e o tipo (atuariais) de informações obtidas.

2288 Essa ressalva é importante, pois a confiabilidade dos resultados da
2289 implementação de procedimentos de formação de classes de equivalência não
2290 deriva exclusivamente dos resultados desta revisão crítica de literatura, mas,
2291 principalmente, do sucesso de um programa de pesquisa que já conta com
2292 quase cinquenta anos de desenvolvimento, marcados predominantemente pelo
2293 uso de delineamentos de sujeito único (ver Sidman, 1994, para uma revisão
2294 histórica da área). Diante dessa vasta história de inovações e replicações, os
2295 resultados apresentados por esta pesquisa representam um pequeno passo que
2296 aponta para a efetividade de uma intervenção em relação ao comportamento de
2297 pessoas em grupo.

2298 Considerando o sucesso do uso de delineamentos de sujeito único como
2299 forma de descrever relações funcionais acerca da efetividade com que
2300 procedimentos de formação de classes entre estímulos desenvolvem
2301 comportamentos relevantes em contextos educacionais, sucesso esse
2302 representado pela longa história dessa área e pelos resultados da presente
2303 pesquisa, é oportuno recomendar que outros pesquisadores da área da
2304 educação adotem mais vezes o uso de delineamentos de sujeito único.
2305 Novamente ressaltamos que essa recomendação não é feita como forma de
2306 combater o uso de delineamentos intersujeitos, uma vez que já reconhecemos o
2307 valor atuarial das informações apresentadas por esse tipo de pesquisa, mas
2308 como forma de fortalecer as evidências acerca das variáveis ambientais
2309 determinantes das mudanças no comportamento das pessoas.

2310 Essa dupla recomendação pode criar contextos que favoreçam uma
2311 maior aproximação entre as comunidades de analistas do comportamento e de
2312 educadores. Essa aproximação poderia reduzir as críticas quanto a estratégias
2313 de ensino analítico-comportamentais (Bernstein & Chase, 2013), principalmente
2314 aquelas derivadas das diferenças culturais entre esses dois grupos, ao fazer com
2315 que ambas as comunidades compartilhem mais práticas, principalmente aquelas
2316 relacionadas à produção e comunicação do conhecimento.

2317 Os resultados desta pesquisa parecem demonstrar que o
2318 desenvolvimento de tecnologias comportamentais, principalmente a EBI, segue
2319 o caminho sugerido por Smith (2013), de que o desenvolvimento e validação de
2320 uma intervenção pode requerer estágios múltiplos, nos quais o uso de
2321 delineamentos intrasujeitos pode cumprir um papel crucial, em etapas iniciais,

2322 para o desenvolvimento e refinamento de intervenções que compõem um
2323 pacote, enquanto delineamentos intersujeitos podem cumprir um papel de
2324 validação do pacote de intervenção como um todo, ao produzir respostas para
2325 importantes questões atuariais, como decidir se deve-se investir na
2326 implementação de uma intervenção ou outra, dependendo da efetividade que
2327 elas apresentam quando aplicadas a grupos amplos.

2328 A partir do esclarecimento dos dois sentidos da questão sobre a
2329 fundamentação de uma prática em evidências e considerando-se a importância
2330 que um modelo de estágios múltiplos para o desenvolvimento e validação de
2331 intervenções, é defensável uma posição de complementariedade entre o uso de
2332 delineamentos inter e intrasujeito. Essa complementariedade poderia facilitar a
2333 colaboração entre analistas do comportamento e a comunidade de educadores,
2334 como sugerimos anteriormente. Contudo, tal proposta não implica abrir mão de
2335 críticas a instituições que sugerem modelos padronizados de integração de
2336 informações (como é o caso da WWC), baseados em critérios de exclusão. Tais
2337 modelos conflitam com as práticas analítico-comportamentais, que favorecem a
2338 integração de informações obtidas durante a história do desenvolvimento de uma
2339 área de pesquisa por meio de replicações.

2340

2341

2342 6 CONCLUSÃO

2343

2344

2345 Com o crescimento e difusão do movimento das PBE no campo médico,
2346 e sua extrapolação para a área da educação, foi criado um cenário no qual
2347 analistas do comportamento se deparam com questionamentos sobre as
2348 evidências que fundamentam as intervenções que desenvolvem. Muitas
2349 estratégias didáticas analítico-comportamentais foram elaboradas e avaliadas,
2350 mas nenhuma delas tornou-se uma tendência dominante na área da educação.
2351 Quanto ao ensino superior, o principal problema responsável por essa rejeição
2352 se deve ao fato de que analistas do comportamento têm preferido usar
2353 repertórios mais simples e fáceis de caracterizar como objetivo de ensino,
2354 enquanto a comunidade de educadores valoriza o ensino de habilidades
2355 complexas, atribuídas a atividades cognitivas superiores. Em resposta a esse
2356 questionamento, propusemos que o uso de procedimentos para a formação de
2357 classes de equivalência entre estímulos **apresenta processos** comportamentais
2358 sofisticados, que outras áreas tendem a chamar de “cognitivos”, para o
2359 desenvolvimento de comportamentos complexos. O uso desses procedimentos
2360 no campo da educação contou, inicialmente, com pesquisas voltadas ao ensino
2361 de repertórios elementares, mas, recentemente, os mesmos procedimentos têm
2362 sido empregados no ensino de repertórios acadêmicos a alunos do ensino
2363 superior.

2364 Diante desse contexto, no qual analistas do comportamento se deparam
2365 com questionamentos acerca das evidências de suas práticas, perguntamos se
2366 o uso de EBI para ensinar repertórios acadêmicos a alunos do ensino superior
2367 contava com evidências suficientes para que fosse recomendada como uma
2368 PBE. Aproveitamos esse questionamento para refletir acerca da noção de
2369 evidência, e, por meio do exame dos compromissos epistemológicos
2370 subjacentes tanto ao movimento das PBE (positivismo), como do behaviorismo
2371 radical (pragmatismo), distinguimos entre o sentido epistemológico e o atuarial
2372 da indagação quanto às evidências que fundamentam uma prática. Esses dois
2373 sentidos têm sido utilizados de maneira indistinta, e, assim, gerado certos
2374 conflitos, mas com o desembaraçar dos significados apontamos que eles não
2375 são necessariamente conflitantes.

2376 Escolhemos os procedimentos para a condução de revisões
2377 sistemáticas da WWC devido à importância dessa instituição para a EBE, e
2378 revisamos a literatura sobre o uso de EBI como estratégia didática no ensino
2379 superior. Após triagem dos resultados, obtivemos 18 artigos cuja avaliação levou
2380 a três pesquisas que “Atendem os Padrões WWC para Delineamentos de
2381 Grupo”. Diante desse resultado, pode-se considerar a EBI apresenta Forte
2382 Evidência acerca de sua efetividade, mas com uma pequena extensão de suas
2383 evidências. A média do tamanho de efeito da EBI é alta o suficiente para
2384 recomendá-la como estratégia didática principal ou complementar em uma
2385 disciplina e estimar que a maioria dos alunos apresentará alto desempenho nos
2386 comportamentos aprendidos. Além disso, os comportamentos ensinados nas
2387 pesquisas consistiam em repertórios complexos, e, dessa forma, contradizem as
2388 críticas acerca da preferência de analistas do comportamento pelo ensino de
2389 repertórios mais simples.

2390 É necessário, contudo, certa cautela diante desses resultados.
2391 Considerando que as duas pesquisas desta revisão que foram avaliadas como
2392 “Atendem aos Padrões WWC” ensinavam comportamentos estatísticos, é
2393 defensável que seria mais adequado apontar que o uso de EBI mostrou-se
2394 efetivo para o ensino de estatística no ensino superior. Diante do contraste entre
2395 essa limitação e a recomendação de que o uso de EBI como estratégia didática
2396 no ensino superior possa ser considerada uma PBE, refletimos acerca dos
2397 parâmetros da WWC para reunir evidências de pesquisas e constatamos que os
2398 problemas de generalidade de resultados, como foi o caso da revisão conduzida
2399 nesta pesquisa, estão relacionados à adoção de critérios excludentes para as
2400 revisões conduzidas pela WWC.

2401 Além do problema da generalidade dos resultados, os critérios da WWC
2402 podem limitar as informações disponibilizadas nas revisões ao descartar a
2403 maioria das pesquisas disponíveis na literatura, uma vez que poucas dentre elas
2404 atingem esses critérios avaliativos. Esta limitação pode ser considerada
2405 prejudicial, na medida em que não disponibiliza para os leitores das revisões
2406 muitas das informações presentes em um grande número de pesquisas. Muitas
2407 das pesquisas excluídas das sínteses de resultados apresentadas nesta revisão
2408 de fato acrescentavam informações novas e importantes quanto ao uso de EBI,
2409 e mesmo que tais pesquisas fossem prejudicadas devido a problemas

2410 metodológicos seria possível relativizar o impacto de tais problemas nos
2411 resultados dessas pesquisas ao se considerar a abrangente história de
2412 pesquisas sobre os ombros das quais os estudos revisados neste trabalho se
2413 apoiam. Diante da importância atribuída por analistas do comportamento à
2414 história de descobertas e replicações, consideramos que a tradição analítico-
2415 comportamental não é compatível com as PBE quanto aos critérios para
2416 agregação dos resultados das pesquisas.

2417 Mesmo diante dessa incompatibilidade, a presente pesquisa pode ser
2418 considerada um exemplo do uso bem-sucedido de delineamentos de grupo por
2419 parte de analistas do comportamento. Esperamos que esse exemplo incentive
2420 mais analistas do comportamento a utilizar tais delineamentos de pesquisa como
2421 forma de mostrar aos pesquisadores, aos profissionais e à população os
2422 impactos estimados das intervenções por eles desenvolvidas.
2423 Concomitantemente, esperamos que a vasta história de pesquisas com
2424 delineamentos de sujeito-único que fundamentam a área de equivalência de
2425 estímulos, e, conseqüentemente, a área de EBI, incentive mais educadores a
2426 utilizar esse tipo de delineamento como forma de descrever relações funcionais
2427 entre suas intervenções e as mudanças nos comportamentos das pessoas. Por
2428 fim, refletimos que uma maior difusão do uso desses delineamentos de pesquisa
2429 nessas duas áreas poderia criar condições mais favoráveis para uma maior
2430 aproximação e comunicação entre elas, na medida em que seus pesquisadores
2431 passariam a apresentar mais práticas em comum, discutindo-as e
2432 aperfeiçoando-as de modo conjunto.

2433

2434

2435 **6.1 Limitações e recomendação para futuras pesquisas**

2436

2437

2438 O ensino superior foi escolhido, nesta revisão, como critério para
2439 avaliação do uso de EBI. Contudo, como argumentamos anteriormente, a
2440 história das pesquisas na área de equivalência de estímulos é mais antiga e
2441 abrangente (Sidman, 1994), e as primeiras aplicações do conhecimento
2442 produzido nessa área foram direcionadas ao desenvolvimento de repertórios
2443 verbais elementares. Seria oportuno avaliar se o uso de procedimentos de

2444 equivalência de estímulos em outros níveis da educação formal pode ser
2445 considerado uma prática baseada em evidências.

2446 É importante notar que a presente pesquisa revisou uma parcela
2447 pequena, ainda que específica, da área de equivalência de estímulos.
2448 Considerando que é possível atribuir a Fields et al. (2009), e Fineup, Critchfield
2449 e Covey (2009), as primeiras pesquisas sobre o uso de EBI no ensino superior,
2450 essa área não possui, ainda, dez anos de desenvolvimento. Dessa forma, a
2451 presente revisão é limitada a uma quantidade que pode ser considerada
2452 pequena de artigos avaliados, e novas revisões, futuramente, deverão ser
2453 conduzidas com mais artigos do que aqueles que aqui dispusemos.

2454 Discutimos, anteriormente, sobre os problemas que a adoção de critérios
2455 excludentes traz para revisões e sumários de resultados, como aqueles
2456 adotados pela WWC, em contraposição à agregação mais abrangente de
2457 resultados de pesquisa utilizados por algumas ciências (Stockard & Wood,
2458 2016). Uma revisão sobre o uso de EBI no ensino superior que analisasse o
2459 mesmo conjunto de textos sob os dois parâmetros poderia oferecer um contraste
2460 mais claro sobre as diferenças entre os sumários de resultados assim produzidos
2461 em relação àqueles apresentados nesta pesquisa, oferecendo assim
2462 fundamento mais consistente para tal argumento.

2463 Os dados acerca da efetividade do uso de EBI no ensino superior foram
2464 criticados quanto à sua generalidade. Isso aponta que são necessárias
2465 pesquisas sobre o uso de EBI para ensinar comportamentos relacionados a
2466 outras disciplinas do ensino superior. Como apresentado anteriormente, a
2467 maioria das pesquisas avaliou o ensino de estatística e neuroanatomia funcional.
2468 É oportuno, portanto, recomendar a realização de pesquisas para aumentar a
2469 abrangência das pesquisas desta área, por meio da avaliação da efetividade
2470 dessa estratégia para ensinar comportamentos relacionados a disciplinas mais
2471 discursivas, como muitas presentes em ciências sociais, artes e filosofia.

2472 O uso de EBI como estratégia didática responde a algumas das críticas
2473 realizadas por educadores quanto a estratégias de ensino propostas por
2474 analistas do comportamento para o ensino superior (Bernstein & Chase, 2013),
2475 principalmente aquelas relacionadas à complexidade dos comportamentos
2476 ensinados. Seria oportuno avaliar as reações de educadores e alunos quanto à
2477 implementação de EBI em um contexto aplicado não apenas em relação ao

2478 desempenho, como também em relação à validação social. Esse tipo de
2479 investigação, como aquela realizada por Greville, Dymond e Newton (2016),
2480 poderia produzir informações importantes sobre a aceitação da comunidade que
2481 recebe uma PBE, potencializando o aperfeiçoamento da intervenção. Não seria
2482 desejável implementar uma prática com boas evidências quanto à sua
2483 efetividade, mas que é aversiva para a comunidade que a receberá. Além disso,
2484 o relato dos sentimentos e opiniões de alunos, educadores e gestores escolares
2485 quanto à implementação da EBI poderia ajudar a identificar a possível existência
2486 de contingências aversivas não detectadas por investigações que se concentram
2487 exclusivamente no desempenho dos alunos.
2488

REFERÊNCIAS

- 2489
- 2490
- 2491 Albright, L., Schnell, L., Reeve, K. F., & Sidener, T. M. (2016). Using stimulus
2492 equivalence-based instruction to teach graduate students in applied
2493 behavior analysis to interpret operant functions of behavior. *Journal of*
2494 *Behavioral Education, 25*(3), 290–309.
- 2495 Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of
2496 applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91–
2497 97.
- 2498 Bernstein, D., & Chase, P. N. (2013). Contributions of behavior analysis to higher
2499 education. In Madden, G. J., Dube, W. V., Hackenberg, T. D., Hanley, G.
2500 P., & Lattal, K. A. (Eds.), *APA handbook of behavior analysis: Translating*
2501 *principles into practice*. (Vol. 2, pp. 523–543). Washington: American
2502 Psychological Association.
- 2503 Boyce, T. E., & Hineline, P. N. (2002). Interteaching: A strategy for enhancing the
2504 user-friendliness of behavioral arrangements in the college classroom. *The*
2505 *Behavior Analyst, 25*, 215–226.
- 2506 Branch, M. N., & Pennypacker, H. S. (2013). Generality and generalization of
2507 research findings. In Madden, G. J., Dube, W. V., Hackenberg, T. D.,
2508 Hanley, G. P., & Lattal, K. A. (Eds.), *APA handbook of behavior analysis:*
2509 *Methods and principles*. (Vol. 1, pp. 151-175). Washington: American
2510 Psychological Association.
- 2511 Burns, M. K. (2012) Meta-analysis of single-case design research: Introduction
2512 to the special issue. *Journal of Behavioral Education, 21*, 175-184.
- 2513 Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental
2514 design for evidence-based practice. *American Journal of Speech-Language*
2515 *Pathology, 21*(4), 397-414.
- 2516 Christophe, M., Elacqua, G., Martinez, M., & Oliveira, J. B. A. (2015). *Educação*
2517 *baseada em evidências: Como saber o que funciona em educação*.
2518 Brasília: Instituto Alfa e Beto.
- 2519 Claridge, J. A., & Fabian, T. C. (2005). History and development of evidence-
2520 based medicine. *World Journal of Surgery, 29*(5), 547-553.
- 2521 Cohen, A. M., Stavri, P. Z., & Hersh, W. R. (2004). A categorization and analysis
2522 of the criticisms of evidence-based medicine. *International Journal of*
2523 *Medical Informatics, 73*(1), 35-43.
- 2524 Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*
2525 (2 ed.). New Jersey: Pearson.
- 2526 Creel, R. E. (1987). Skinner on science. In Modgil, S. & Modgil, C. (Eds.), *B. F.*
2527 *Skinner: Consensus and controversy* (pp. 107-114). Philadelphia: Falmer
2528 Press.

- 2529 Davidoff, F., Haynes, B., Sackett, D., & Smith, R. (1995). Evidence based
2530 medicine. *BMJ: British Medical Journal*, 310(6987), 1085.
- 2531 Davies, P. (1999). What is evidence-based education?. *British Journal of*
2532 *Educational Studies*, 47(2), 108-121.
- 2533 Davies, P. (2007). Revisões sistemáticas e a Campbell Collaboration. In Thomas,
2534 G., & Pring, R. (Eds.), *Educação baseada em evidências: A utilização dos*
2535 *achados científicos para a qualificação da prática pedagógica*. Porto Alegre:
2536 ARTMED.
- 2537 De Rose, J. C. C. (1993). Classes de estímulos: Implicações para uma análise
2538 comportamental da cognição. *Psicologia: Teoria e Pesquisa*, 9(2), 283–
2539 303.
- 2540 Detrich, R., & Lewis, T. (2013). A decade of evidence-based education: Where
2541 are we and where do we need to go?. *Journal of Positive Behavior*
2542 *Interventions*, 15(4), 214-220.
- 2543 Djulbegovic, B., Guyatt, G. H., & Ashcroft, R. E. (2009). Epistemologic inquiries
2544 in evidence-based medicine. *Cancer Control*, 16(2), 158-168.
- 2545 Dutra, L. H. A. (2005). *Oposições filosóficas: A epistemologia e suas polêmicas*.
2546 Florianópolis: Editora da UFSC.
- 2547 Dutra, L. H. A. (2010). *Introdução à epistemologia*. São Paulo: Editora UNESP.
- 2548 Elliot, J. (2007). Tornando educativa a prática baseada em evidências. In
2549 Thomas, G., & Pring, R. (Eds.), *Educação baseada em evidências: A*
2550 *utilização dos achados científicos para a qualificação da prática pedagógica*
2551 (pp. 177-200). Porto Alegre: ARTMED.
- 2552 Evidence-Based Medicine Working Group. (1992). Evidence-based medicine: A
2553 new approach to teaching the practice of medicine. *Jama*, 268(17), 2420.
- 2554 Fields, L., Travis, R., Roy, D., Yadlovker, E., Aguiar-Rocha, L. D., & Sturmey, P.
2555 (2009). Equivalence class formation: A method for teaching statistical
2556 interactions. *Journal of Applied Behavior Analysis*, 42(3), 575-593.
- 2557 Fienup, D. M., & Critchfield, T. S. (2010). Efficiently establishing concepts of
2558 inferential statistics and hypothesis decision making through contextually
2559 controlled equivalence classes. *Journal of Applied Behavior Analysis*, 43(3),
2560 437–462.
- 2561 Fienup, D. M., & Critchfield, T. S. (2011). Transportability of equivalence-based
2562 programmed instruction: Efficacy and efficiency in a college
2563 classroom. *Journal of Applied Behavior Analysis*, 44(3), 435-450.
- 2564 Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain-behavior
2565 relations economically with stimulus equivalence technology. *Journal of*
2566 *Applied Behavior Analysis*, 43(1), 19–33.

- 2567 Fienup, D. M., Mylan, S. E., Brodsky, J., & Pytte, C. (2016). From the laboratory
2568 to the classroom: The effects of equivalence-based instruction on
2569 neuroanatomy competencies. *Journal of Behavioral Education*, 25(2), 143-
2570 165.
- 2571 Goldenberg, M. J. (2006). On evidence and evidence-based medicine: Lessons
2572 from the philosophy of science. *Social Science & Medicine*, 62(11), 2621-
2573 2632.
- 2574 Greville, W. J., Dymond, S., & Newton, P. M. (2016). The student experience of
2575 applied equivalence-based instruction for neuroanatomy teaching. *Journal*
2576 *of educational evaluation for health professions*, 13.
- 2577 Hammersley, M. (2007). Algumas questões sobre a prática baseada em
2578 evidências na educação. In Thomas, G., & Pring, R. (Eds.), *Educação*
2579 *baseada em evidências: A utilização dos achados científicos para a*
2580 *qualificação da prática pedagógica*. Porto Alegre: ARTMED.
- 2581 Higgins, J. P., & Green, S. (Eds.) (2011). *Cochrane handbook for systematic*
2582 *reviews of interventions* (Vol. 4). New Jersey: John Wiley & Sons.
- 2583 Hjørland, B. (2011). Evidence-based practice: An analysis based on the
2584 philosophy of science. *Journal of the Association for Information Science*
2585 *and Technology*, 62(7), 1301-1310.
- 2586 Hodkinson, P. & Smith, J. K. (2007) A relação entre pesquisa, políticas e prática.
2587 In Thomas, G., & Pring, R. (Eds.), *Educação baseada em evidências: A*
2588 *utilização dos achados científicos para a qualificação da prática pedagógica*
2589 (pp. 161-176). Porto Alegre: ARTMED.
- 2590 Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012).
2591 Considerations for the systematic analysis and use of single-case
2592 research. *Education and Treatment of Children*, 35(2), 269-290.
- 2593 Iversen, I. H. (2013). Single-case research methods: An overview. In Madden, G.
2594 J., Dube, W. V., Hackenberg, T. D., Hanley, G. P., & Lattal, K. A. (Eds.),
2595 *APA handbook of behavior analysis: Methods and principles*. Washington:
2596 American Psychological Association.
- 2597 Keller, F. S. (1968). Goodbye, teacher.... *Journal of Applied Behavior*
2598 *Analysis*, 1(1), 79-89.
- 2599 Kienen, N., Kubo, O. M., & Botomé, S. P. (2013). Ensino programado e
2600 programação de condições para o desenvolvimento de comportamentos:
2601 Alguns aspectos no desenvolvimento de um campo de atuação do
2602 psicólogo. *Acta Comportamental*, 21(4), 481-494.
- 2603 Lattal, K. A. & Chase, P. N. (2003) *Behavior theory and philosophy*. Boston:
2604 Springer.

- 2605 Lattal, K. A., & Laipple, J. S. (2003). Pragmatism and behavior analysis. In K. A.
2606 Latal & P. N. Chase (Eds.), *Behavior theory and philosophy* (pp. 41-61).
2607 Boston: Springer.
- 2608 Leonardi, J. L. (2016). *Prática baseada em evidências em psicologia e a eficácia*
2609 *da análise do comportamento clínica*. Tese de Doutorado, Universidade de
2610 São Paulo.
- 2611 Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of*
2612 *Applied Behavior Analysis*, 25, 51–57.
- 2613 Lorena, A. B., Castro-Caneguim, J. F., & Carmo, J. S. (2013). Habilidades
2614 numéricas básicas: Algumas contribuições da análise do comportamento.
2615 *Estudos de Psicologia*, 18(3), 439–446.
- 2616 Matos, M. A. (1999). Controle de estímulo condicional, formação de classes
2617 conceituais e comportamentos cognitivos. *Revista Brasileira de Terapia*
2618 *Comportamental Cognitiva*, 1(2), 159–178.
- 2619
2620 Modgil, S. & Modgil, C. (1987). *B. F. Skinner: Consensus and controversy*. New
2621 York: Falmer Press.
- 2622 Moore, J. (2008). *Conceptual foundations of radical behaviorism*. New York:
2623 Sloan.
- 2624 Moreira, M. B. (2004). Em casa de ferreiro, espeto de pau: O ensino de análise
2625 experimental do comportamento. *Revista Brasileira de Terapia*
2626 *Comportamental e Cognitiva*, 6(1), 73-79.
- 2627 Perone, M., & Hursh, D. E. (2013). Single-case experimental designs. In Madden,
2628 G. J., Dube, W. V., Hackenberg, T. D., Hanley, G. P., & Lattal, K. A. (Eds.),
2629 *APA Handbook of behavior analysis, Vol. 1: Methods and principles*.
2630 Washington: American Psychological Association.
- 2631 Rabelo, L. Z., & De Rose, J. C. (2015). É possível fazer uma análise
2632 comportamental da inteligência? *Revista Brasileira de Análise Do*
2633 *Comportamento*, 11(1), 80–92.
- 2634 Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack,
2635 B. (2004). What counts as evidence in evidence-based practice? *Journal of*
2636 *advanced nursing*, 47(1), 81-90.
- 2637 Sackett, D. L. (1997). Evidence-based medicine. *Seminars in perinatology*, 21(1),
2638 3-5.
- 2639 Schnaitter, R. (1987). Knowledge as action: The epistemology of radical
2640 behaviorism. In Modgil, S. & Modgil, C. (Eds.), *B. F. Skinner: Consensus*
2641 *and controversy* (pp. 57-68) . New York: Falmer Press.
- 2642 Scholten, R. J. P. M., Clarke, M., & Hetherington, J. (2005). The Cochrane
2643 Collaboration. *European Journal of Clinical Nutrition*, 59, 147-149.

- 2644 Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). *Metodologia*
2645 *de pesquisa em psicologia* (R. C. Costa, Trad). Porto Alegre: AMGH.
- 2646 Sidman, M. (1994). *Equivalence relations and behavior: A research story*.
2647 Boston: Authors Cooperative.
- 2648 Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to
2649 sample: An expansion of the testing paradigm. *Journal of the Experimental*
2650 *Analysis of Behavior*, 37(1), 5–22.
- 2651 Skinner, B. F. (1945). The operational analysis of psychological
2652 terms. *Psychological Review*, 52(5), 270.
- 2653 Skinner, B. F. (1953). *Science and human behavior*. New York: Pearson.
- 2654 Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-
2655 Crofts.
- 2656 Skinner, B. F. (1974). *About behaviorism*. New York: Alfred A. Knopf.
- 2657 Slavin, R. E. (2002). Evidence-based education policies: Transforming
2658 educational practice and research. *Educational Researcher*, 31(7), 15-21.
- 2659 Slavin, R. E. (2008). Perspectives on evidence-based research in education —
2660 What works? Issues in synthesizing educational program
2661 evaluations. *Educational Researcher*, 37(1), 5-14.
- 2662 Smith, J. D. (2012). Single-case experimental designs: A systematic review of
2663 published research and current standards. *Psychological Methods*, 17(4),
2664 510.
- 2665 Smith, T. (2013). What is evidence-based behavior analysis?. *The Behavior*
2666 *Analyst*, 36(1), 7-33.
- 2667 Souza, D. G., De Rose, J. C. C., Calcagno, S., & Galvão, O. de F. (2004). Análise
2668 comportamental da aprendizagem de leitura e escrita e a construção de um
2669 currículo suplementar. In M. M. C. Hübner & M. Marinotti (Eds.), *Análise do*
2670 *comportamento para a educação: Contribuições recentes*. (Vol. 1). Santo
2671 André: ESETec.
- 2672 Stockard, J., & Wood, T. W. (2017). The threshold and inclusive approaches to
2673 determining “best available evidence”: An empirical analysis. *American*
2674 *Journal of Evaluation*, 38(4), 471-492.
- 2675 Thomas, G., & Pring, T. (2007). *Educação baseada em evidências: A utilização*
2676 *dos achados científicos para a qualificação da prática pedagógica* (R. C.
2677 Costa, Trans.). Porto Alegre: Artmed.
- 2678 Tourinho, E. Z. (1996). Behaviorismo radical, representacionismo e
2679 pragmatismo. *Temas em Psicologia*, 4(2), 41-56.

- 2680 Tourinho, E. Z., & Neno, S. (2003). Effectiveness as truth criterion in behavior
2681 analysis. *Behavior and Philosophy*, 31, 63-80.
- 2682 What Works Clearinghouse. (2017a). *What Works Clearinghouse procedures*
2683 *handbook* (4.0). Washington, DC: US Department of Education.
- 2684 What Works Clearinghouse. (2017b). *What Works Clearinghouse standards*
2685 *handbook* (4.0). Washington, DC: US Department of Education.
- 2686 Zinn, T. E., Newland, M. C., & Ritchie, K. E. (2015). The efficiency and efficacy
2687 of equivalence-based learning: A randomized controlled trial. *Journal of*
2688 *Applied Behavior Analysis*, 48(4), 865-882.
- 2689 Zuriff, G. E. (1980). Radical behaviorist epistemology. *Psychological*
2690 *Bulletin*, 87(2), 337-350.