

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Faculdade de Filosofia e Ciências Humanas

Programa de Pós-graduação em Psicologia: Cognição e Comportamento

Daniel Ezequiel Pinto

**REVISÃO DE TREINOS AUTOMATIZADOS BASEADOS NA RFT E ALTERAÇÕES
METODOLÓGICAS NO TREINO SMART**

Belo Horizonte

2021

Daniel Ezequiel Pinto

**REVISÃO DE TREINOS AUTOMATIZADOS BASEADOS NA RFT E ALTERAÇÕES
METODOLÓGICAS NO TREINO SMART**

Dissertação de mestrado apresentada à Faculdade de Filosofia e Ciência Humanas da Universidade Federal de Minas Gerais como parte dos requisitos de obtenção do grau de Mestre em Psicologia: Cognição e Comportamento.

Área de concentração: Neuropsicologia do Desenvolvimento

Orientadora: Profa. Dra. Thais Porlan de Oliveira

Belo Horizonte

2021

153.4 P659r 2021	<p>Pinto, Daniel Ezequiel. Revisão de treinos automatizados baseados na RFT e alterações metodológicas no treino SMART [manuscrito] / Daniel Ezequiel Pinto. - 2021. 98 f. Orientadora: Thais Porlan de Oliveira.</p> <p>Dissertação (mestrado) - Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas. Inclui bibliografia.</p> <p>1. Psicologia – Teses. 2. Comportamento - Teses. I. Oliveira, Thais Porlan de. II. Universidade Federal de Minas Gerais. Faculdade de Filosofia e Ciências Humanas. III. Título.</p>
------------------------	---

Ficha catalográfica elaborada por Vilma Carvalho de Souza - Bibliotecária - CRB-6/1390



ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO **DANIEL EZEQUIEL PINTO**

Realizou-se, no dia 30 de novembro de 2021, às 09:00 horas, Videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *Revisão de treinos automatizados baseados na RFT e alterações metodológicas no treino SMART*, apresentada por DANIEL EZEQUIEL PINTO, número de registro 2019668720, graduado no curso de PSICOLOGIA, como requisito parcial para a obtenção do grau de Mestre em PSICOLOGIA: COGNIÇÃO E COMPORTEAMENTO, à seguinte Comissão Examinadora: Prof(a). Thais Porlan de Oliveira - Orientador (UFMG), Prof(a). João Henrique de Almeida (Universidade Estadual de Londrina), Prof(a). Viviane Verdu Rico (Universidade Federal de Minas Gerais).

A Comissão considerou a dissertação:

Aprovada

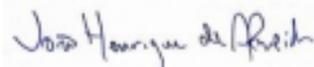
Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

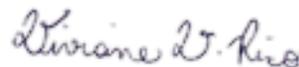
Belo Horizonte, 30 de novembro de 2021.


THAIS PORLAN
DE
OLIVEIRA:25537
919842

Prof(a). Thais Porlan de Oliveira (Doutora)



Prof(a). João Henrique de Almeida (Doutor)



Prof(a). Viviane Verdu Rico (Doutora)

AGRADECIMENTOS

Agradeço primeiramente a minha esposa, Iara Lúcio, por ter me dado apoio e carinho em todas as etapas desse processo. Sem você, seria impossível conseguir terminar esse mestrado. A pandemia e o desgoverno trouxeram muito sofrimento durante esse período, mas nossa relação me permitiu passar por esses anos de uma forma muito mais amena. Agradeço a minha família, por ter aberto todas as portas ao longo da vida, para que eu pudesse estudar e me dedicar a uma área que eu amo. Como bom analista do comportamento, não acredito em meritocracia. Sem o contexto de onde eu vim, eu não poderia ter chegado aonde cheguei. Obrigado, família. Obrigado ao meu grande amigo, Carlos Alexandre Lima Melo por ter me incentivado a tentar esse mestrado e ter me acompanhado no início desse processo. Agradeço a minha orientadora, Profa. Thais Porlan, por ter aceitado trabalhar ao meu lado nesse tema tão recente e pelas orientações ao longo desse processo. Agradeço aos meus amigos da pós, Fernanda Machado, Larissa Coutens e Átila Cedro por terem segurado minha mão em todas as etapas, rindo e chorando, ao longo dessa trajetória. Agradeço aos meus amigos de infância e programadores, Norton Pigozzo e Vinícius Garcia, pelas aulas, reuniões e sessões de programação (e desabafo) em conjunto. Agradeço também a todos os demais professores do departamento, em especial a professora Viviane Verdu, pelas conversas e incentivo prestado, mas também ao professor Edson Huziwara, que teve grande papel na minha paixão pela análise do comportamento ainda na graduação e ao professor Maycoln Teodoro, pelos anos de iniciação científica. Obrigado ao professor Antonio Jaeger e ao colega Pedro Saulo pelas ajudas nas análises estatísticas. Obrigado ao professor João Henrique de Almeida por ter aceitado o convite a banca e pelas contribuições prestadas na defesa. Obrigado às estudantes de iniciação científica, Vanessa Torchetti e Diuliana Nader pelas ajudas nas coletas e interesse no meu trabalho. Por último, agradeço a coordenação pedagógica e aos alunos da Escola Municipal Anne Frank e a todos os pais e crianças que aceitaram contribuir na participação desse experimento. Obrigado a todos que me auxiliaram nessa jornada, a lista de nomes completa teria o tamanho dessa dissertação.

RESUMO

O SMART é um protocolo automatizado de treino baseado na Teoria das Molduras Relacionais (RFT) que permite aos participantes desenvolverem o responder relacional derivado (DRR). Além dele, há outros treinos baseados na RFT para desenvolver o DRR. Dois estudos foram realizados para explorar aspectos teóricos e empíricos de alternativas metodológicas desses protocolos de treino. O primeiro estudo consistiu em um artigo de revisão sistemática sobre os protocolos de treinos baseados na RFT, incluindo o SMART, objetivando comparar variáveis metodológicas e desempenho dos participantes no SMART e em outros treinos automatizados baseados na RFT que desenvolvem o DRR. Os artigos foram selecionados a partir de duas técnicas, a busca por palavras-chave e o uso de *snowballing*. Após a etapa de seleção, os 27 estudos experimentais foram comparados de acordo com a amostra; desempenho e critérios de maestria em testes de implicação combinatória (IC); uso de testes psicométricos; protocolo, estrutura e duração dos treinos; estímulos utilizados e respostas relacionais; e uso do treino de múltiplos exemplares. Os resultados da revisão sugeriram uma heterogeneidade entre os artigos encontrados que dificultou a comparação entre eles e foram destacadas uma série de variáveis independentes que podem ser manipuladas em futuros experimentos. O segundo estudo consistiu na aplicação de um jogo de computador baseado no SMART, com quatro estágios. Participaram 40 crianças entre 9 e 12 anos que foram divididas em quatro grupos experimentais de acordo com a inserção de múltiplos conjuntos de estímulos no estágio dois, estágio três, em ambos ou nenhum deles. Foram analisados os escores dos participantes nos quatro estágios e a probabilidade de acerto nos diferentes tipos de tentativas do primeiro estágio de acordo com a distância nodal da questão, tipo de relação (igualdade a/ou oposição) e direção da questão presente nas tentativas. Os resultados demonstraram que não houve diferenças significativas entre os grupos com relação ao uso de múltiplos conjuntos de estímulos. O estágio três demonstrou-se mais fácil que os demais em todos os grupos. Os tipos de tentativas demonstraram diferenças significativas a partir da probabilidade de acerto de acordo com a distância nodal e relação da questão, porém a linearidade da questão não impactou nas chances de acerto. Os resultados corroboram com a hipótese que a ordenação do treino SMART não é eficiente em termos de aumento gradual de dificuldade. Esse aspecto pode impactar no desempenho de alguns participantes no treino e pode ser crítico se pensarmos no uso do SMART para outras populações, como crianças com desenvolvimento atípico. Os achados dos dois

trabalhos, o artigo de revisão e o artigo empírico, oferecem sugestões para pesquisadores interessados em estudos a partir de treinos automatizados baseados na RFT.

Palavras-chave: responder relacional derivado, treino SMART, Teoria das Molduras Relacionais, treino de múltiplos exemplares, treinos computadorizados.

ABSTRACT

SMART is an automated training protocol based on relational frame theory (RFT) that allows participants to develop derived relational responding (DRR). Besides SMART there are other RFT-based training protocols that aim the development of DRR. Two studies were carried out to explore theoretical and empirical aspects of methodological alternatives to these training protocols. The first study consisted of a systematic review on RFT-based training protocols, including SMART, aiming to compare methodological variables and participants' performance on SMART and those other automated RFT-based training protocols that develop DRR. The articles were selected based on two techniques, the search for keywords and the use of snowballing. After the selection step, the 27 experimental studies were compared according to the sample; performance and mastery criteria in combinatorial entailment (CE) tests; use of psychometric tests; training protocol, structure and duration; stimuli used and relational responses; and use of multiple-exemplar training. The results of the review suggested a heterogeneity among the articles found, which made the comparison between them difficult. A series of independent variables that can be manipulated in future experiments were highlighted. The second study consisted of the application of a computer game based on SMART, with four stages. Forty children between 9 and 12 years old participated, and were splitted into four experimental groups according to the insertion of multiple sets of stimuli in stage two, stage three, in both or none of them. Participants' scores in the four stages and the probability of success in the different types of tasks in the first stage were analyzed according to the nodal distance, type of relationship (sameness or opposition) and direction of the questions present in the tasks. The results showed that there were no significant differences between groups regarding the use of multiple sets of stimuli. Stage three proved to be easier than the others in all groups. The types of tasks showed significant differences based on the probability of success according to the nodal distance and the relation of the question, but the linearity of the question did not impact the chances of getting it right. The results support the hypothesis that the SMART training order is not efficient in terms of gradual increase in difficulty. This aspect can impact the performance of some participants in training and can be critical if we consider the use of SMART for other populations, such as children with atypical development. The findings of the two studies, the review article and the empirical article, offer suggestions for researchers interested in studies based on automated training based on RFT, highlighting the SMART protocol.

Keywords: derived relational responding, SMART training, relational frame theory, multiple exemplar training, computerized training

LISTA DE FIGURAS

Estudo 1

Figura 1 - Fluxograma do método <i>snowballing</i>	24
Figura 2 - Fluxograma dos resultados gerais com as técnicas de busca em base de dados e <i>snowballing</i>	26
Figura 3 - Número de artigos publicados por faixa etária dos participantes.....	28

Estudo 2

Figura 1 -Tela de exemplo do estágio final do jogo utilizado na coleta de dados, adaptado a partir do SMART.....	60
Figura 2 - ANCOVA comparando escore dos quatro grupos no estágio 4.....	63
Figura 3 - Probabilidade de acerto a partir da interação entre relação da questão e tipo de tentativa	68

LISTA DE TABELAS

Estudo 1

Tabela 1 - Descritores utilizados em cada base de dados.....	25
Tabela 2 - Instrumentos, referência, molduras relacionais treinadas e amostra.....	27

Estudo 2

Tabela 1 - Três exemplos de tentativas utilizadas nos estágios 27, 28 e 29 do SMART.....	52
Tabela 2 - Todas as doze possibilidades de questão dos três estágios do SMART (27,28 e 29)...	55
Tabela 3 - Divisão dos grupos experimentais de acordo com a apresentação dos estímulos (MET)	62
Tabela 4 - Escore médio de cada grupo por estágio.....	63
Tabela 5 - Resultados das Correlações de Pearson.....	65
Tabela 6 - Tipos de tentativas deduzidas a partir dos parâmetros.....	66
Tabela 7 - Post hoc de Bonferroni comparando tipos de tentativas.....	67
Tabela 8 - Post hoc de Bonferroni comparando a interação entre tipos de tentativa e relação da questão.....	67

SUMÁRIO

Introdução geral	14
Estudo 1	18
Método	23
Resultados	26
SMART	29
REP	31
RCP	34
TARPA	36
T-IRAP	40
Discussão	41
Conclusão	46
Referências	47
Estudo 2	51
Método	58
Participantes	58
Instrumentos	58
Procedimentos	60
Análise dos dados e resultados	63
Discussão	69
Conclusão	73
Conclusão geral	74
Referências	76
Anexo A	78
Anexo B	81

INTRODUÇÃO

Em 2011 foi desenvolvido um treino automatizado (Cassidy et al., 2011) baseado na Teoria das Molduras Relacionais (*Relational Frame Theory* [RFT]; S. C. Hayes et al., 2001). O propósito desse treino era de aumentar o repertório dos participantes no responder relacional derivado, o principal operante estudado pela RFT, a partir de tentativas envolvendo molduras relacionais de igualdade, oposição e comparação. Desde seu desenvolvimento o treino, chamado SMART (*strengthening mental abilities with relational training*), foi aplicado em mais de 300 participantes (Amd & Roche, 2018; Cassidy et al., 2016; J. Hayes & Stewart, 2016; McLoughlin et al., 2020, 2021; Thirus et al., 2016) e segundo os pesquisadores demonstrou resultados superiores ao treino mais popular de memória de trabalho (McLoughlin et al., 2020). Dentre as evidências da utilização desse treino estão o aumento no escore de testes de habilidades acadêmicas e em testes de inteligência padronizados (McLoughlin et al., 2021).

As evidências sobre o SMART são consistentes com as hipóteses levantadas por autores proponentes da RFT (S. C. Hayes et al., 2001). Para estes, o responder relacional derivado (DRR) é um comportamento basilar para o desenvolvimento de repertórios mais complexos como a linguagem, o raciocínio lógico e abstrato. Em resumo, o DRR é um comportamento operante que se refere a como os indivíduos são capazes de derivar novas relações entre estímulos sem que elas sejam treinadas diretamente. Por exemplo, se lhe é dito que A é maior que B, e B é maior que C, é possível inferir, ou seja, derivar, que A é maior que C, mesmo que a relação entre A e C nunca tenha sido explicitada diretamente.

Para a RFT o DRR possui três propriedades. A implicação mútua (Se A é maior que B, então implica-se que B é menor que A), a implicação combinatória (Se é maior que B e B é maior que C, implica-se que A é maior que C) e a transformação de função. Para a compreensão desta última propriedade pode ser interessante imaginar, por exemplo, a marca da sua bebida favorita. Se a bebida possui uma função apetitiva, é esperado que a marca adquira a mesma função. Pode ser que diante da marca você salive, ainda que não esteja diante da bebida propriamente dita. Ou seja, a marca, um símbolo ou nome, teve sua função transformada a partir da relação estabelecida de igualdade entre marca e a bebida. Nesse caso, a relação condicionada foi de igualdade, enquanto no exemplo anterior foram utilizadas relações de comparação (“maior que” e “menor que”). O DRR possibilita uma série de relações entre estímulos (Hayes et al., 2001).

A natureza das relações entre estímulos estabelecidas por meio do DRR é treinada a partir de estímulos providos pela comunidade verbal ao longo da vida de um indivíduo. Uma criança, ainda em desenvolvimento da sua linguagem, pode ouvir coisas como: “isso é um cachorro”, diante do animal pela primeira vez. Em outras ocasiões, diante de estímulos distintos, pode ouvir “isso é um gato”, “isso é um peixe”, “isso é um pato”. Diante de múltiplos exemplos similares, a criança então aprende que a dica verbal “*isso é*” denota uma relação de igualdade entre qualquer estímulo apontado (o animal) e a palavra atribuída ao estímulo (“gato”, “cachorro”, etc.). A relação então é explicitada pela dica provida pelo contexto, daí o nome utilizado pela RFT para esse estímulo de “dica contextual”. Outras dicas contextuais indicam outras relações entre os estímulos, tais como de oposição (*A é oposto a B*, i.e. *claro é oposto de escuro*, *fundo é oposto de raso*), temporais (*A foi antes de B*, i.e. *ontem foi antes de hoje*), espaciais (*A está acima de B*, i.e. *o nono andar está acima do oitavo*), etc (Hayes et al., 2001).

As relações entre os estímulos podem também ser treinadas de forma arbitrária ou não-arbitrária. Por exemplo, no Brasil uma moeda de 25 centavos é maior em termos de dimensões físicas do que uma moeda de 50 centavos. Essa é uma relação não-arbitrária, treinada a partir de características objetivas como a diferença visível do diâmetro das moedas. Todavia, pode-se dizer que uma moeda de 50 centavos é maior que uma moeda de 25 centavos quando se considera o valor da mesma e ignora-se a dimensão física. Esse valor foi arbitrariamente estabelecido pela comunidade verbal, afinal, tanto os numerais quanto o sistema financeiro são convenções culturais e simbólicas. Dessa forma, pode-se observar que a dica verbal sobre cada relação (neste exemplo, *maior que*) precisam estar contextualizadas para que o responder relacional seja emitido de modo preciso de acordo com a convenção simbólica estabelecida (Hayes et al., 2001).

O SMART (Cassidy et al., 2016) procura, a partir da exposição do indivíduo a múltiplos exemplares de relações e de estímulos, desenvolver o repertório do DRR no participante que passa pelo treino. O treino do SMART é realizado a partir de uma série de tentativas silogísticas com pseudopalavras, utilizadas como estímulos abstratos. Por exemplo, em uma tentativa apresentam-se duas premissas em sequência e dicas contextuais indicando as relações entre essas pseudopalavras: “Xug é igual a Jor” e “Jor é igual a Mip”. Em seguida é apresentada uma pergunta como “Xug é igual a Mip?”. O participante então deve responder a pergunta de acordo com as relações apresentadas, respondendo “sim” ou “não” (a resposta é “sim”). As tentativas então variam de acordo com a complexidade, na medida em que o participante avança no treino,

tornando-se progressivamente mais difíceis. O procedimento é completamente automatizado e mediado por um software.

Estudos com o SMART foram aplicados em uma série de contextos com amostras de participantes entre seis e 18 anos, tendo demonstrado aumento dos escores em diferentes testes de inteligência como WASI e WISC (Cassidy et al., 2016; J. Hayes & Stewart, 2016), Raven (Amd & Roche, 2018) e K-BIT (McLoughlin et al., 2021). Também foram realizados estudos similares com crianças com dificuldades acadêmicas (Cassidy & Roche, 2011) e com crianças em vulnerabilidade socioeconômica em Bangladesh (Amd & Roche, 2018). Ainda que preliminares, os resultados a partir do SMART indicam a potencialidade desse treino no desenvolvimento do DRR e os impactos positivos do SMART demonstrados psicometricamente. Todavia, o SMART não é o único protocolo de treino automatizado baseado na RFT. Outros protocolos vêm sendo estudados empiricamente utilizando softwares para mediação dos experimentos, com objetivos similares. Para citar alguns, *Training and Assessment of Relational Precursors and Abilities* (TARPA; Moran et al., 2010), *Relational Completion Procedure* (Dymond et al., 2007), *Relational Evaluation Procedure* (Cullinan et al., 2001) e *Teaching Implicit Relational Assessment Procedure* (T-IRAP; Kilroe et al., 2014), a serem apresentados posteriormente. Todos estudam e buscam desenvolver o mesmo comportamento operante, o DRR.

Apesar das diversas terminologias e tecnologias de treino, consideramos que, desde que diferentes protocolos sejam desenvolvidos e aplicados visando a compreensão e o ensino do mesmo comportamento operante, suas evidências podem ser estudadas em conjunto. Essas evidências, agrupadas, permitirão que pesquisadores interessados no DRR agreguem descobertas e componentes metodológicos de diferentes estudos, com protocolos automatizados, às suas agendas de pesquisa, ou até mesmo utilizem a síntese desses protocolos na criação de uma possível nova tecnologia para treino de repertório, fundamentada na RFT.

Portanto, visando auxiliar o desenvolvimento de treinos baseados no responder relacional e o aprofundamento teórico de estudos empíricos fundamentados na RFT, tivemos como objetivo geral neste trabalho identificar variáveis que impactam nos resultados do treino SMART. Para tanto, conduzimos dois estudos. O primeiro consistiu em uma revisão de literatura que buscou comparar procedimentos do SMART com outros modelos de treinos automatizados ou mediados por software, também baseados na RFT. O segundo estudo, de natureza empírica, avaliou o desempenho de participantes entre 9 e 12 anos em tentativas experimentais realizadas online,

seguindo os mesmos parâmetros de três estágios do SMART a partir de relações de igualdade e oposição. As crianças foram divididas em grupos que realizavam tentativas a partir de conjuntos iguais de estímulos entre os estágios e grupos que realizavam tentativas com estímulos variando a cada exercício. Essas variações foram elaboradas de acordo com as evidências encontradas em outros treinos baseados no treino de múltiplos exemplares (May et al., 2017; Walsh et al., 2014). Avaliamos se houve diferenças no escore dos participantes de cada grupo, considerando os dados socioeconômicos. Também realizamos análises avaliando os critérios que possuíam maior probabilidade de acerto entre tentativas de um mesmo estágio.

ESTUDO 1

Uma revisão de literatura dos protocolos automatizados baseados na RFT

O uso de ferramentas automatizadas de ensino são uma ambição de analistas do comportamento desde a década de 60 com a criação das máquinas de aprendizagem Skinnerianas (Holland, 1960; Skinner, 1958). Afinal, a automação de experimentos e procedimentos de ensino permite uma série de benefícios como a mediação imediata e consistente de reforços, além de consistência na obtenção de dados sem necessidade de aumento do número de experimentadores (Kilroe et al., 2014).

Em 2011, Cassidy et al. publicaram um trabalho no qual foi apresentado um treino automatizado baseado na Teoria das Molduras Relacionais (*Relational Frame Theory* ou *RFT*; S.C. Hayes et al., 2001). No protocolo, os participantes eram treinados a responder perguntas envolvendo o operante chamado *responder relacional derivado* (DRR). Após o treino, sete de oito crianças do estudo demonstraram aumentos de ao menos um desvio padrão na escala Wechsler de inteligência, com os participantes saindo da faixa de QI abaixo da média para um QI médio ou acima da média (Cassidy et al., 2011). Os ganhos se mantiveram mesmo quatro anos após o estudo (Cassidy et al., 2013). O treino proposto foi posteriormente adaptado e recebeu o nome de *Strengthening Mental Abilities With Relational Training* (SMART; Colbert et al., 2018). Desde então, as pesquisas a partir do SMART continuam acumulando evidências em uma série de estudos com diferentes testes de inteligência e escalas psicométricas (Amd & Roche, 2018; Cassidy et al., 2016; J. Hayes & Stewart, 2016; McLoughlin et al., 2018, 2020, 2021; Thirus et al., 2016).

O DRR, operante treinado pelo SMART, também foi relacionado a resultados em testes de inteligência em outros estudos (O’Hora et al., 2005, 2008; O’Toole & Barnes-Holmes, 2009) e o SMART chegou a demonstrar resultados superiores ao treino mais popular de memória de trabalho (McLoughlin et al., 2021). O impacto de treinos em escores de inteligência ou testes psicométricos é, no entanto, um tópico controverso mesmo dentre cientistas do comportamento que apontam que os relatos de aumentos nos escores podem significar erros de método ou de testagem dos participantes (Magiati & Howlin, 2001; Reed et al., 2005). Contudo, as evidências do SMART, ainda que preliminares, apontam a possibilidade de estudos futuros com treinos automatizados baseados na RFT que comprovem o impacto desses treinos sobre habilidades cognitivas (Roche et al., 2013; McLoughlin et al., 2021).

Além do SMART, há outros protocolos automatizados, mediados por *software*, que buscam o desenvolvimento do DRR. Dentre eles podemos citar o *Relational Evaluation Procedure* (REP; Cullinan et al., 2001), o *Relational Completion Procedure* (RCP; Dymond et al., 2007), o *Training and Assessment of Relational Precursors and Abilities* (TARPA; Moran et al., 2010) e o *Teaching Implicit Relational Assessment Procedure* (T-IRAP; Kilroe et al., 2014). Cada uma dessas tecnologias possui suas respectivas peculiaridades de formato, mas todos os protocolos se conformam a variações de uma estrutura metodológica similar. Todos esses instrumentos envolvem a divisão entre fase treino, com *feedback* sobre erros ou acertos, e fases teste, sem *feedback*. Os componentes de cada um dos protocolos são os mesmos presentes em qualquer exemplo de DRR (S.C. Hayes et al., 2001), a saber: a implicação mútua; a implicação combinatória; a transformação de função; a dica contextual e os estímulos relacionados. Diferentes protocolos utilizam diferentes estímulos no treino, dicas contextuais e opções de resposta, podendo variar entre estímulos abstratos, como símbolos (e.g. REP; Stewart et al., 2004) e estímulos não-abstratos relacionados por propriedades físicas, como figuras similares (e.g. RCP; Dymond et al., 2007).

No SMART são apresentadas aos participantes perguntas com premissas relacionando pseudopalavras e é exigido deles que pressionem “sim” ou “não” de acordo com as relações e questões apresentadas, em até 30 segundos. Uma típica tentativa do SMART com duas premissas e uma questão sobre elas é exemplificada a seguir:

XUG é maior que JAC.

JAC é maior que GUP.

XUG é maior que GUP?

Ao longo de 55 estágios as tentativas variam em níveis de complexidade segundo o número de premissas apresentado, as relações presentes nas premissas, a direção das premissas, as relações presentes nas questões, a direção da questão e a distância nodal entre os estímulos presentes nas questões (Amd & Roche, 2018). Cada estágio do SMART é dividido em duas etapas, uma etapa teste e uma etapa treino, sendo necessários 16 acertos consecutivos (100%) para avanço entre as duas etapas. Em caso de falha, na etapa teste ou no treino, o participante reinicia a etapa treino (com *feedback*) do mesmo estágio até que se consiga 16 acertos consecutivos. Após os 16 acertos, avança-se para a etapa teste do mesmo estágio. Após mais 16 acertos consecutivos na etapa de teste, avança-se para o próximo estágio, com um novo grau de complexidade. Ao errar, o

participante reinicia outra tentativa dentro do mesmo estágio com um conjunto completamente novo de estímulos. O conjunto de estímulos (i.e. pseudopalavras) nunca se repete entre uma tentativa e outra.

No protocolo automatizado chamado de REP há três variações de formato principais. Uma variação baseada em procedimento *go/no-go* (e.g. Smeets et al., 2004) e duas que utilizam opções de resposta para os participantes (e.g. Cullinan et al., 2001; Stewart et al., 2004). Na variação baseada em *go/no-go* os participantes são requisitados a pressionar a barra de espaço mediante apresentações de pares de estímulos pertencentes ao mesmo conjunto (e.g. Smeets et al., 2004). Dois estímulos são dispostos um após o outro no centro da tela, a tela é limpa, e o participante possui até 5 segundos para pressionar ou não a barra de acordo com o conjunto apresentado. É esperado que o participante pressione a barra 50% das vezes, já que em metade das apresentações as tentativas serão de estímulos não pertencentes ao mesmo conjunto e na outra metade (pressão a barra) são apresentados estímulos pertencentes ao mesmo conjunto.

Já no procedimento que utiliza a variação baseada em opções de resposta um estímulo é apresentado no centro da tela por um segundo, a tela é limpa, e posteriormente outro estímulo é apresentado na tela por um segundo (e.g. Cullinan et al., 2001). Em seguida, são apresentadas duas opções de resposta que permanecem na tela, até a resposta do participante. Em sequência o programa emite o feedback de acordo com a resposta esperada. Em um formato um pouco diferente, no estudo de Stewart et al. (2004), duas opções de resposta permanecem na tela enquanto os estímulos e a relação entre eles ficam dispostos em quadrados localizados no centro e parte inferior da tela (a complexidade do procedimento não cabe nesse trabalho, para mais detalhes ver Stewart et al., 2004).

No procedimento do RCP é apresentada aos participantes uma tela dividida em duas partes, uma superior na qual encontra-se uma relação incompleta e uma inferior, que dispõe diferentes opções de estímulos (Dymond et al., 2007). A tentativa dos participantes, então, é selecionar um estímulo dentre as opções e arrastá-lo até os espaços em branco que ficam situados ao lado da dica contextual, no intuito de completar a relação requerida. Por exemplo: na parte superior encontra-se “A é maior que ___”. Enquanto na parte inferior encontram-se os estímulos B, C e D e o participante escolhe qual é adequada para completar a lacuna. Após a construção da resposta, duas opções de botões são apresentadas na tela: um botão escrito “comece novamente”, que faz com que todos os estímulos voltem para a posição inicial permitindo ao participante uma nova chance

de montar sua resposta, e outro botão escrito “terminar a tentativa”, que avança para a próxima questão.

No TARPA é apresentada uma tela com um estímulo modelo no centro e estímulos de comparação nos cantos e, para pontuar, o participante deve escolher a resposta correta clicando no estímulo apropriado em cada uma das tentativas (Moran et al., 2014). No estágio 1 são testadas respostas de discriminação simples nas quais o participante escolhe o estímulo a ser reforçado no estágio, dentre uma lista de estímulos. O estágio 1 possui três níveis de dificuldade que variam de acordo com a quantidade de estímulos a serem comparados. No estágio 2 são respostas de discriminações não-arbitrárias, baseadas em similaridade física, também com três níveis de dificuldade que variam de acordo com o número de estímulos. Já o estágio 3 trata de discriminações arbitrárias, de forma que as relações desse estágio são previamente treinadas.

Por fim, no protocolo T-IRAP pares de estímulos são apresentados simultaneamente na tela e é requisitado ao participante que pressione uma de duas teclas no computador para a escolha da resposta correta de acordo com a relação a ser reforçada. Em seguida é apresentado o feedback, a tela é limpa e uma nova tentativa é apresentada (Kilroe et al., 2014).

Tendo em vista a existência de diferentes protocolos de treino automatizados desenvolvidos a partir do SMART surge a questão se há possibilidade de comparação entre os diferentes protocolos, e em que termos isso seria possível. Pressupõe-se que, por estudarem o mesmo operante (o DRR) sob a mesma perspectiva teórica da RFT, os protocolos automatizados possam ser comparados entre si. Entretanto, não encontramos dados na literatura que apontem a comparação entre esses treinos e quais parâmetros seriam mais adequados para essa comparação. Assim, tivemos como objetivo no presente trabalho realizar uma revisão sistemática da literatura para comparar variáveis metodológicas e desempenho dos participantes no SMART e em outros treinos automatizados baseados na RFT para desenvolver o DRR.

Para este estudo escolhemos alguns parâmetros de comparação baseados na literatura sobre variáveis que impactam em treinos relacionais (Arntzen, 2012; Fields et al., 2020). Assim, consideramos: as molduras relacionais treinadas; a caracterização da amostra; o desempenho e critérios de maestria em testes de implicação combinatória (IC); o uso de testes psicométricos; o protocolo, estrutura e duração dos treinos; os estímulos utilizados e respostas relacionais exigidas; além do uso de treino de múltiplos exemplares (MET; ver S.C. Hayes et al., 2001).

Há evidências que diferentes protocolos e estruturas de treino aumentam ou diminuem a porcentagem de participantes que demonstram formações de redes relacionais, mais especificamente, classes de equivalência (Arntzen, 2012). A estrutura do treino pode variar entre linear, um-para-muitos ou muitos-para-um. Em uma estrutura linear as tentativas de treino são distribuídas de tal forma que em um treino para formação de duas classes de estímulos (A1-B1-C1-D1 e A2-B2-C2-D2) os estímulos e suas comparações são apresentados de forma ordenada em sequência ($A \rightarrow B$ e posteriormente $B \rightarrow C$). Já em treinos de um-para-muitos o mesmo modelo é utilizado em uma série de comparações ($A \rightarrow B, A \rightarrow C, A \rightarrow D$), enquanto em treinos de muitos-para-um diversos estímulos são utilizados como modelo enquanto a comparação não varia ($B \rightarrow A, C \rightarrow A, D \rightarrow A$) (Arntzen, 2012).

Além do impacto que as estruturas de treino podem ter sobre a formação de redes relacionais de participantes, há diferenças na porcentagem de participantes que completam os treinos em protocolos simples-para-complexo em relação a protocolos de apresentação simultânea (Smeets et al., 2006). Em protocolos simples-para-complexo tentativas de implicação combinatória são testadas prévia e separadamente de tentativas que testam simultaneamente a implicação mútua e a implicação combinatória (e.g. treina-se primeiramente as relações $A \rightarrow B$ e $B \rightarrow C$. Em sequência, testa-se primeiramente a implicação $A \rightarrow C$ e somente depois a implicação $C \rightarrow A$). Outras variáveis que a literatura aponta como relevantes são o tamanho das redes relacionais e o tipo de estímulos utilizados, que também podem impactar nos resultados dos treinos (Fields et al., 1999; Finn et al., 2018; Moran et al., 2014).

Por último, a utilização do treino de múltiplos exemplares foi avaliada neste estudo a partir de dois critérios. Primeiramente foi avaliado se os autores utilizam o termo "múltiplos exemplares" em algum momento no texto. Em segundo lugar, como definido por Cassidy et al. (2011), avaliamos se os participantes precisavam emitir uma resposta relacional comum a partir de diferentes conjuntos de exemplares de estímulos. As definições de treino de múltiplos exemplares variam entre estudos (Holth, 2017; LaFrance & Tarbox, 2019), porém a definição de Cassidy foi escolhida, posto que nossa base de comparação é o treino SMART (Cassidy et al., 2016).

MÉTODO

Procedimentos e análise dos dados

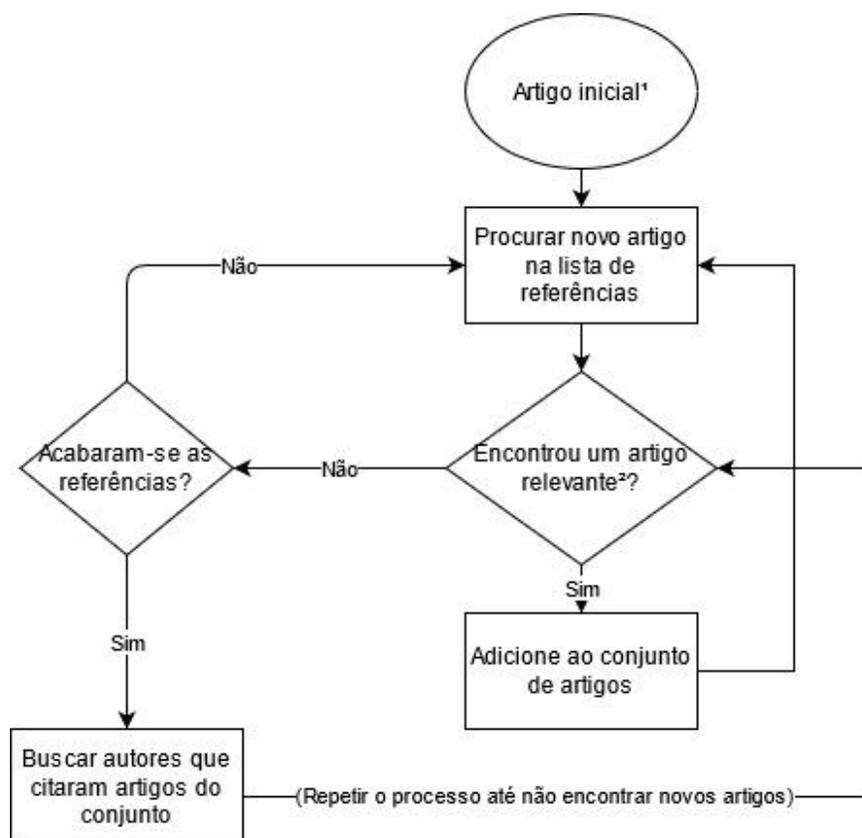
Utilizamos o método PRISMA, desenvolvido para ajudar revisores de literatura com critérios de como reportar o processo de revisão sistemática e seus resultados (Page et al., 2020). Organizamos os critérios de seleção dos artigos de acordo com o método SPIDER (Cooke et al., 2012). Esse método sugere que a seleção de artigos a partir de busca de bases de dados seja sistematizada segundo a amostra, o fenômeno de interesse, o desenho experimental, a forma de avaliação dos dados e o tipo de pesquisa. A amostra selecionada incluiu estudos com seres humanos de todas as faixas etárias. O fenômeno de interesse foi o DRR. Quanto ao desenho experimental, os artigos selecionados deveriam utilizar protocolos de treino automatizados e precisavam avaliar o desempenho dos participantes em tentativas de responder relacional. Incluímos exclusivamente estudos empíricos.

Nossa pesquisa abarcou artigos de 2001, quando o trabalho seminal da RFT (S. C. Hayes et al., 2001) foi publicado, até março de 2021. Reunimos as publicações em uma única biblioteca através do software Zotero. Utilizamos duas técnicas de revisão em cada etapa da coleta de artigos, sendo elas o *snowballing* (Wohlin, 2014) e a busca sistemática em bases de dados.

O *snowballing* (traduzido como “fazer uma bola de neve”) é uma técnica de revisão que consiste em, a partir de um conjunto de artigos, verificar quais referências são citadas (*backward snowballing*, “para trás”) ou verificar quais autores citam os artigos inseridos no conjunto (*forward snowballing*, “para frente”). Em nosso trabalho optamos por iniciar o *snowballing* por meio da busca na ferramenta Google Acadêmico de artigos mais recentes, publicados em 2020 e 2021, do principal protocolo de treino aqui analisado, utilizando o nome completo do instrumento SMART e entre aspas para garantir o uso do termo exato (“*strengthening mental abilities with relational training*”). A partir daí, seguiu-se com a leitura de todos os títulos e resumos das referências relevantes citadas nos trabalhos publicados em 2020 e 2021, até não haver mais artigos relacionados ao objetivo, publicados entre 2001 e 2021. O fluxo do método *snowballing* está descrito na Figura 1.

Figura 1

Fluxograma do método snowballing



Nota. ¹foi utilizado o artigo de McLoughlin et al. (2021) como artigo inicial por ser o artigo mais recente encontrado com o uso do SMART. ²consideramos como artigos relevantes àqueles que seguiam os critérios de seleção definidos anteriormente através do método SPIDER (Cooke et al., 2012).

Ao se chegar ao final do fluxo da técnica de *backward snowballing*, os artigos de saída deram origem a um novo conjunto de estudos. Esse novo conjunto foi utilizado para o *forward snowballing* (“para frente”), ou seja, buscamos autores que citaram o conjunto de artigos coletados. O processo foi repetido até que não encontrássemos novos artigos dentro dos critérios de inclusão. No processo de *backward snowballing* optou-se por incluir quaisquer artigos empíricos com protocolos de treinos da RFT, mesmo que não automatizados, para que a abrangência da busca fosse a mais completa possível nessa etapa. Já no processo de *forward snowballing* restringimos os artigos selecionados exclusivamente a artigos que treinavam molduras relacionais ou componentes específicos do DRR de forma automatizada, via software, atendendo mais especificamente ao objetivo deste trabalho. Quando havia dúvida sobre a inclusão a partir dos títulos e resumos, eram lidas a seção de métodos e a de instrumentos.

A segunda técnica utilizada no método envolveu o emprego de palavras-chave e busca sistemática em bases de dados a partir das palavras "*relational frame theory*", "*derived relational responding*" e "*training*". Limitamos a busca a artigos publicados entre 2001 e março de 2021. Realizamos as buscas em cinco bases de dados: Pubmed; Psycinfo; Scopus; Web of Science e BVS Salud. Os estudos selecionados a partir das buscas em base de dados passaram pelo crivo do pesquisador principal e dois auxiliares de pesquisa treinados, utilizando a ferramenta Rayyan (Ouzzani et al., 2016). Os termos das buscas utilizados em cada base são expostos na Tabela 1.

Tabela 1

Descritores utilizados em cada base de dados.

Base de dados	Termos de busca/descriptores utilizados
Pubmed	((<i>relational frame theory</i>) AND (<i>training</i>) AND (<i>derived relational responding</i>))
Psycinfo	Any Field: <i>relational frame theory</i> AND Any Field: <i>training</i> AND Any Field: <i>derived relational responding</i>
Scopus	TITLE-ABS-KEY (<i>relational frame theory</i>) AND TITLE-ABS-KEY (<i>training</i>) AND TITLE-ABS-KEY (<i>derived relational responding</i>))
Web of science	<i>relational frame theory</i> (All fields) and <i>training</i> (All Fields) and <i>derived relational responding</i> (All fields)
BVS Salud	título/resumo/assunto (<i>relational frame theory</i>) AND (<i>training</i>) AND (<i>derived relational responding</i>)

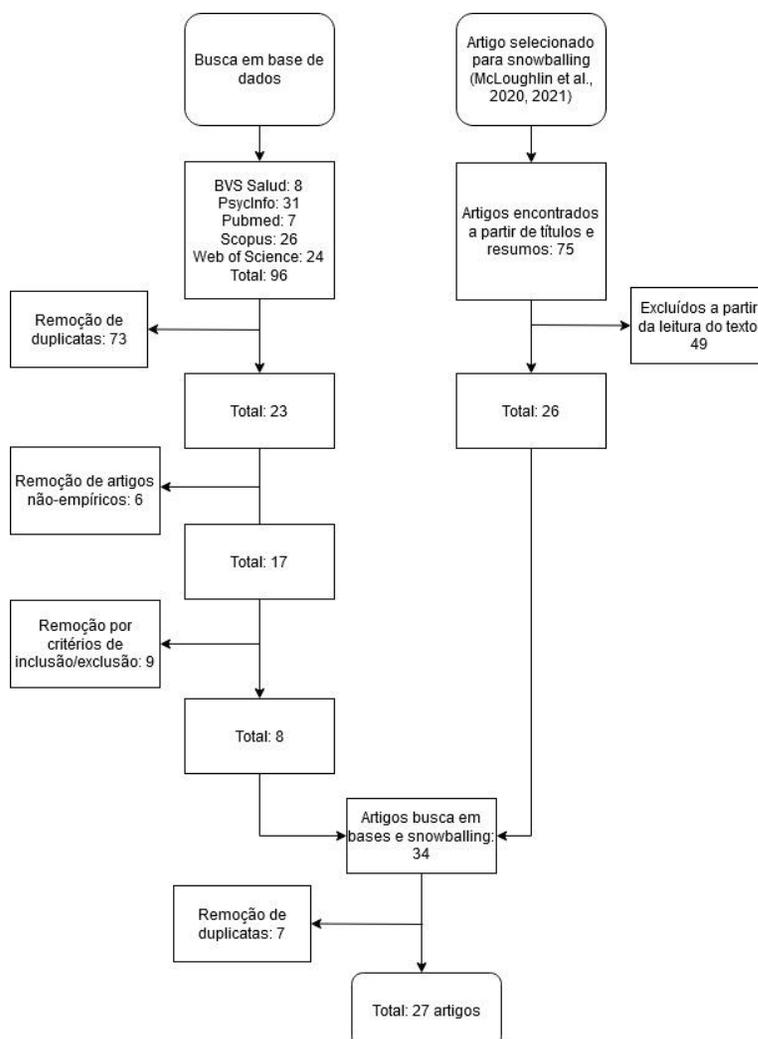
Após a etapa de seleção dos artigos eles foram comparados considerando: as molduras relacionais utilizadas, a amostra; desempenho e critérios de maestria em testes de implicação combinatória (IC); uso de testes psicométricos; protocolo, estrutura e duração dos treinos; estímulos utilizados e respostas relacionais; e uso do MET.

RESULTADOS

Os dados gerais e fluxo das buscas utilizando as duas técnicas escolhidas são explicitados no fluxograma na Figura 2.

Figura 1

Fluxograma dos resultados gerais com as técnicas de busca em base de dados e snowballing.



Encontramos ao todo cinco instrumentos, incluindo o SMART, a saber: SMART, REP, RCP, TARPA e T-IRAP. Os artigos selecionados, diferentes molduras relacionais treinadas e dados da amostra, por instrumento, são apresentados na Tabela 2. Dentre as molduras relacionais treinadas encontramos artigos com molduras de igualdade, diferença, oposição, comparação (maior/menor) e molduras espaciais. A maioria dos artigos utilizou treinos a partir de molduras de igualdade (24 artigos) e oposição (12). Os demais artigos utilizaram molduras de comparação (11)

e diferença (5). Somente um dentre os 27 artigos utilizou molduras relacionais espaciais, a partir de uma versão adaptada do RCP (May et al., 2017). Destacamos que os trabalhos com o TARPA se restringiram a molduras de igualdade.

Tabela 2

Instrumentos, referência, molduras relacionais treinadas e amostra.

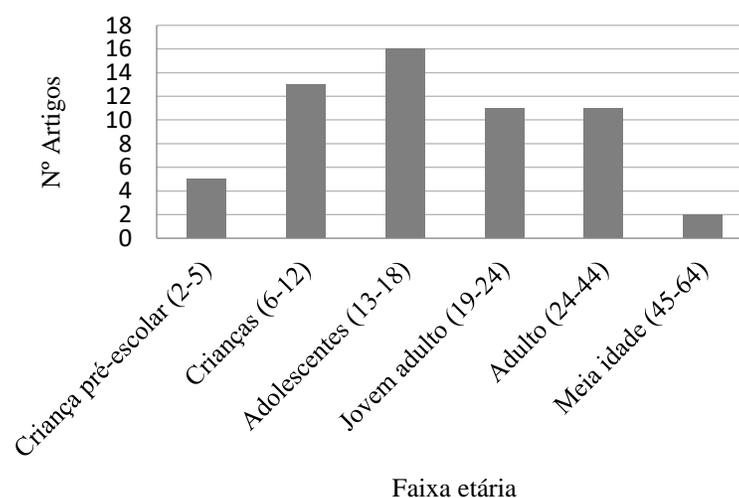
Instrumentos	Referência	Molduras	Nº de participantes	Faixa etária	Diagnóstico
SMART	Cassidy et al., 2016	Igualdade, oposição, comparação.	45	11-17	-
	J. Hayes & Stewart, 2016	Igualdade, oposição, comparação.	28	10-11	-
	Thirus et al., 2016	Igualdade, oposição, comparação.	21	16-18	-
	Colbert et al., 2018	Igualdade, oposição, comparação	26	15-17	-
	Amd & Roche, 2018	Igualdade, oposição, comparação.	35	6-14	-
	McLoughlin et al., 2018	Igualdade, oposição, comparação	8	22-52	-
	McLoughlin et al., 2020	Igualdade, oposição, comparação.	175	12-15	-
	McLoughlin et al., 2021	Igualdade, oposição, comparação.	49	6-10	-
REP	Cullinan et al., 2001	Igualdade, diferença.	28	16-30	-
	Stewart et al., 2004	Igualdade, diferença.	5	19-31	-
	Smeets et al., 2004	Igualdade.	27	18-31	-
	Smeets et al., 2006	Igualdade.	52	18-54	-
	J. Hayes et al., 2016	Igualdade, diferença.	26	2-5	-
	Corbett et al., 2017	Igualdade, diferença.	9	7-11	TEA
RCP	Dymond et al., 2007	Igualdade, oposição.	12	21-38	-
	Dymond et al., 2008.	Igualdade, oposição.	15	22-44	-
	Dymond & Whelan, 2010	Igualdade, oposição.	32	18-35	-

	Dymond et al., 2013	Igualdade, oposição.	48	18-30	-
	Munnelly et al., 2013	Comparação.	24	18-34	-
	Walsh et al., 2014	Igualdade.	17	5-18	TEA
	May et al., 2017	Espacial.	17	18-59	-
	Munnelly et al., 2019	Comparação.	29	19-36	-
	Moran et al., 2010	Igualdade.	5	6-13	TEA
TARPA	Kishita et al., 2013	Igualdade.	7	2-6	-
	Moran et al., 2014	Igualdade.	23	2-12	TEA
	Moran et al., 2015	Igualdade.	35	3-15	TEA
T-IRAP	Kilroe et al., 2014	Igualdade, diferença, oposição, comparação.	4	8-10	TEA

Quase metade da amostra total de participantes dos trabalhos seleccionados são provenientes do SMART, 387 dos 802 participantes totais. A caracterização da amostra por idade demonstrou grande variedade, com participantes de dois até 59 anos. Sua estratificação e número de artigos com cada faixa etária é descrita na Figura 3. Alguns artigos possuem uma faixa que se sobrepõe em mais de um estrato. Os participantes dos estudos foram principalmente pessoas neurotípicas, sendo somente seis dos 27 artigos (Kilroe et al., 2014; Moran et al., 2010, 2014, 2015; Walsh et al., 2014) realizados com 93 participantes diagnosticados com transtorno do espectro autista (TEA).

Figura 3

Número de artigos publicados por faixa etária dos participantes



A seguir apresentamos os principais resultados por instrumento relacionados à desempenho e critérios de maestria em testes de IC, uso de testes psicométricos, protocolo, estrutura e duração dos treinos, estímulos utilizados e respostas relacionais e uso do MET.

SMART

Desempenho e critérios de maestria em testes de IC

O número de participantes que passaram em testes de IC não foi reportado nos estudos com o SMART, mas todos os participantes que passaram por pelo menos um estágio do SMART completaram ao menos um teste de IC, a partir de critérios de maestria de 100%. O estudo de Amd & Roche (2018) foi o único que relatou participantes que não passaram no primeiro estágio enquanto demais estudos não relataram quantos estágios os participantes avançaram individualmente.

Uso de testes psicométricos

Todos os estudos encontrados com o SMART avaliaram os participantes a partir de alguma escala psicométrica. Todos os oito estudos avaliaram a inteligência e três estudos (Cassidy et al., 2016; J.Hayes & Stewart, 2016; McLoughlin et al., 2021) utilizaram testes para habilidades acadêmicas. Também foi avaliada a velocidade de processamento (McLoughlin et al., 2020) e resultado em testes de personalidade (McLoughlin et al., 2020, 2021).

Seis dos oito estudos demonstraram ganhos significativos nos testes de inteligência em medidas antes e depois dos treinos (Amd & Roche, 2018; Cassidy et al., 2016; Colbert et al., 2018; J. Hayes & Stewart, 2016; McLoughlin, 2020, 2021). Amd e Roche (2018) utilizaram o teste Raven de inteligência para avaliação do pré e pós-teste e demonstraram, através de uma análise de regressão múltipla, que ganhos no teste no pós-treino foram previstos a partir do último estágios completado pelos participantes ($\beta = .691$, $p < .001$).

Cassidy et al. (2016) avaliaram a inteligência através da escala Wechsler de inteligência para crianças (WISC) e aplicaram um teste t para avaliar os impactos do treino. O aumento no escore de QI foi significativo $t(14) = 19,183$ ($p < 0,001$), com tamanho de efeito grande, d de Cohen = 2,329 (Cohen, 1988).

J. Hayes e Stewart (2016) avaliaram a inteligência através da escala de inteligência Wechsler abreviada (WASI). Tanto o grupo controle ativo quanto experimental tiveram ganhos significativos nos escores de inteligência no pré e pós-intervenção. Todavia, foi encontrada uma diferença significativa entre os dois grupos, com o grupo SMART demonstrando resultados superiores ao grupo controle a partir de um teste t independente comparando os dois grupos no pré e pós-treino $t(26) = 4,303$; $p < 0,001$. Colbert et al. (2018) também utilizaram a escala WASI e os participantes demonstraram aumentos nas duas condições experimentais (experimental e controle), mas um teste t independente a partir de mudanças no QI da linha de base para o pós-treino encontrou uma diferença estatisticamente significativa entre os grupos; $t(24) = 11,87$; $p < 0,001$.

McLoughlin et al. (2018, 2020, 2021) avaliaram a inteligência através da escala breve de inteligência Kaufman. Enquanto no estudo de 2018 não foram encontradas diferenças significativas entre o QI em relação ao grupo controle e o experimental, nos outros dois estudos foram encontradas diferenças significativas no QI não-verbal a partir da interação condição (grupo controle e experimental) x tempo (pré e pós-intervenção) (McLoughlin et al., 2020, 2021).

Entre os três estudos que testaram habilidades acadêmicas nos participantes pós-treino SMART, dois demonstraram ganhos (Cassidy et al., 2016; J. Hayes & Stewart, 2016). McLoughlin et al. (2021) também utilizaram testes de avaliação de habilidades acadêmicas, porém não utilizaram uma comparação pré e pós-intervenção. No estudo de Cassidy et al. (2016) foi realizado um teste t para comparar resultados do pré e pós-treino no teste de habilidades acadêmicas Drumcondra, com os resultados significativos: $t(26) = 5,207$; $p < 0,001$; com tamanho de efeito médio a partir do d de Cohen (0,73). J. Hayes e Stewart (2016) avaliaram habilidades acadêmicas a partir de dois testes: o teste Wechsler, que avaliou soletração, operações matemáticas e habilidades de leitura e o teste Drumcondra, de matemática. As diferenças entre grupos experimental e controle no pré e pós teste foram significativas em todos esses testes.

A velocidade de processamento foi avaliada em McLoughlin et al. (2020) a partir de uma tentativa adaptada pelos autores. Os resultados demonstraram que os dois grupos, experimental e controle ativos, obtiveram maiores resultados na velocidade de processamento no pós-teste.

McLoughlin et al. (2020, 2021) avaliaram a personalidade dos participantes através de dois testes baseado no modelo dos cinco fatores de personalidade. Eles demonstraram que aspectos da personalidade estiveram inversamente relacionados ao número de estágios completados pelos participantes em relação a faceta de volatilidade do traço de neuroticismo, e a polidez do traço de

amabilidade (McLoughlin et al., 2020) além da susceptibilidade a emoções negativas (McLoughlin et al., 2021).

Protocolo, estrutura e duração dos treinos

Nenhum dos estudos com SMART reportou uma categoria de protocolo ou estrutura de treino. Cinco estudos reportaram a duração das sessões e o número de sessões (Amd & Roche, 2018; Cassidy et al., 2016; Colbert et al., 2018; J. Hayes & Stewart, 2016; McLoughlin et al., 2018). O número de sessões variou de no mínimo 5 (McLoughlin et al., 2018) a 222 sessões (Cassidy et al., 2016). As durações de sessões variaram de 30 (Amd & Roche, 2018) a 60 minutos (Cassidy et al., 2016).

Estímulos utilizados e respostas relacionais

Todos os trabalhos utilizaram pseudopalavras como estímulos para o treino e as opções "sim" e "não" para resposta. As dicas contextuais são explícitas na tela ("é igual a", "é oposto a", "é maior que", "é menor que").

Uso do MET

Todos os artigos com SMART utilizaram múltiplos exemplares, variando o conjunto de estímulos a cada tentativa.

REP

Desempenho e critérios de maestria em testes de IC

Quatro dos seis estudos utilizaram testes de IC (Cullinan et al., 2001; Smeets et al., 2004, 2006; Stewart et al., 2004). Dois estudos não realizaram testes de IC (Corbett et al., 2016; J. Hayes et al., 2016) e, por isso, excluímos 35 participantes (2-11 anos) do total da análise. Dos 113 participantes (16-54 anos) restantes a partir dos demais estudos, 68 (60%) atingiram os critérios de maestria.

Oitenta participantes foram expostos ao REP a partir de um formato *go/no-go*, dos quais 47 (58%) passaram no teste de implicação combinatória (Smeets et al., 2004, 2006). Nesse formato, os melhores resultados foram obtidos a partir da inserção de um protocolo simples-para-complexo, da adaptação das instruções e com a inserção do feedback somente em ocasiões de pressão a barra. As adaptações das instruções incluíram requisitar do participante a leitura em voz alta das instruções e a adição da instrução do participante pressionar a barra em 50% das vezes.

Considerando a inserção de todas essas adaptações 15 de 16 (93%) participantes completaram o treino.

Trinta e três participantes foram expostos ao REP a partir de um formato com opções de resposta (Cullinan et al., 2001; Stewart et al., 2004) dos quais 21 (63%) passaram no teste de implicação combinatória. Nesses estudos os melhores resultados foram obtidos em condições experimentais onde foram utilizadas opções de resposta não-arbitrárias de "igual" ou "diferente" ou a partir de opções de resposta arbitrárias treinadas a partir de MET. Nessas condições 100% (17) dos participantes completaram o treino.

Os critérios de maestria no teste de IC variaram entre 80% e 100% de acertos. Três dos quatro artigos utilizaram critérios de 80% (Cullinan et al., 2001; Smeets et al., 2004, 2006) e um utilizou 100% (Stewart et al., 2004).

Uso de testes psicométricos

Somente dois estudos entre os seis avaliaram a correlação entre testes psicométricos e desempenho de crianças no REP (Corbett et al., 2016; J.Hayes et al., 2016), ambos utilizando a escala de linguagem pré-escolar PLS-4. Eles demonstraram correlações significativas a partir de uma análise de correlação de Spearman entre o escore total no PLS-4 e na versão do REP utilizada ($\rho = 0,82$ e $p < 0,05$ em Corbett et al. [2016]; $\rho = 0,87$ e $p < 0,001$ no estudo de J. Hayes et al., [2016]). J. Hayes et al. (2016) também demonstraram a correlação com o escore total de uma escala de inteligência abreviada, o SB5-ABIQ ($\rho = 0,82$; $p < 0,001$).

Protocolo, estrutura e duração dos treinos

Dois artigos especificaram a categoria do protocolo de treino que utilizaram (Smeets et al., 2004; Smeets et al., 2006). Nesses trabalhos eles inseriram um protocolo de treino simples-para-complexo e observaram superioridade no número de participantes que completavam o teste de implicação combinatória em relação aos protocolos simultâneos.

A estrutura do treino só foi categorizada em um dos artigos (Smeets et al., 2006), que comparou uma estrutura linear com uma estrutura muitos-para-um. Os autores não demonstraram diferença entre desempenho dos participantes comparando o grupo com treino linear em relação ao grupo com treino muitos-para-um, já que em ambos os grupos três dos quatro participantes em cada grupo atingiram critérios de implicação combinatória.

Somente um dos estudos não relatou o número de sessões para o treino (Cullinan et al., 2001), reportando somente a duração do treino (entre 20 e 120 minutos). Nos demais estudos o

número de sessões variou entre um e 12. Dois dos seis estudos realizaram um maior número de sessões, mas com sessões curtas de 2 a 12 minutos aproximadamente (Corbett et al., 2016, J. Hayes et al., 2016) e possivelmente devido a faixa etária e diagnóstico dos participantes (participantes de 2 a 5 anos em Corbett et al., 2016; participantes com TEA de 7 a 11 anos em J. Hayes et al., 2016). Três estudos utilizaram menos sessões (uma a três) com durações mais longas, entre 20 e 180 minutos (Smeets et al., 2004, 2006; Stewart et al., 2004).

Estímulos utilizados e respostas relacionais

Todos os estímulos utilizados eram visuais, podendo ser pseudopalavras, formas abstratas ou imagens não-abstratas. Quatro dos seis artigos utilizaram pseudopalavras como estímulos abstratos (Cullinan et al., 2001; Smeets et al., 2004, 2006; Stewart et al., 2004). Dentre esses quatro, um também utilizou formas abstratas combinadas às pseudopalavras (Stewart et al., 2004). Os demais artigos comparavam diversos pares de imagens não-arbitrárias (e.g. 20 pares de fotos de animais) (Corbett et al., 2016; J. Hayes et al., 2016).

Quanto às respostas exigidas dos participantes, três dos seis artigos utilizaram opções de resposta não-arbitrárias (e.g. palavras como "igual", "diferente", "sim" e "não") (Corbett et al., 2016; Cullinan et al., 2001; J. Hayes et al., 2016). Dois utilizaram opções de resposta arbitrárias (símbolos e formas abstratas) (Cullinan et al., 2001; Stewart et al., 2004). Dois artigos utilizaram a pressão a barra de espaços no teclado do computador como opção de resposta, no procedimento chamado *go/no-go* (Smeets et al., 2004, 2006).

Apenas em um dos seis artigos as opções de resposta foram avaliadas experimentalmente (Cullinan et al., 2001). Foram demonstrados desempenhos superiores a partir de opções de resposta "igual" e "diferente" em relação aos grupos com opções de resposta "sim" e "não" e "vai com" (*goes with*) ou "não vai com" (*does not go with*). O desempenho dos participantes das condições experimentais com opções "igual" e "diferente" foi similar ao desempenho nas condições em que eram utilizadas opções de resposta arbitrárias (símbolos), desde que houvesse um pré-treino a partir de estímulos não-abstratos. O artigo de Stewart et al. (2004) também treinou estímulos abstratos como opções de resposta a partir de estímulos não-abstratos.

Uso de MET

Dentre os seis estudos, dois não variaram os conjuntos de estímulos entre tentativas em nenhum momento (Smeets et al., 2004, 2006) e, portanto, não utilizaram um procedimento de MET. Três variaram os exemplares a partir do conjunto de estímulos durante a fase de treino

(Corbett et al., 2016; J.Hayes et al., 2016; Stewart et al., 2004). Dois utilizaram MET na fase de pré-treino, onde eram treinadas opções de resposta arbitrárias (Cullinan et al., 2001; Stewart et al., 2004).

RCP

Desempenho e critérios de maestria em testes de IC

Somente 147 participantes realizaram testes de IC a partir do RCP e foram considerados na análise. Desses, 127 (86%) atingiram os critérios de IC.

Dymond e Whelan (2010) consideraram que o uso de uma resposta confirmatória, ou seja, o participante confirmar sua escolha a partir de duas opções, "finalizar" ou "começar de novo", antes de avançar para a próxima tentativa, aumenta a eficácia do RCP. Nas condições experimentais onde a resposta confirmatória foi utilizada, 7 em 8 (87%) participantes atingiram critérios de maestria no IC em comparação com 1 em 8 (12%) participantes da condição sem resposta confirmatória (Dymond & Whelan, 2010). Todos os demais estudos do RCP empregaram a resposta confirmatória em todas as condições experimentais.

Uma porcentagem menor de participantes completou testes de IC nas tentativas de molduras relacionais espaciais (10 em 17 participantes, 56%) (May et al., 2017). Subtraindo-se da análise as condições experimentais onde o RCP demonstrou-se inferior, ou seja, os 17 participantes de tentativas de molduras espaciais (May et al., 2017) e os 8 participantes da condição sem resposta confirmatória (Dymond & Whelan, 2010), 116 em 122 participantes atingiram critérios de maestria a partir do RCP (95%).

O critério de maestria em todos os estudos variou entre 75% e 90% de acertos. Dois artigos utilizaram critérios de maestria de 75% (Munnelly et al., 2013, 2019), quatro utilizaram critérios de maestria de 87,5% (Dymond et al., 2007, 2008, 2013; Dymond & Whelan, 2010) e dois utilizaram critérios de maestria de 90% (May et al., 2017; Walsh et al., 2014).

Uso de testes psicométricos

Nenhum dos artigos do RCP comparou o desempenho dos participantes ao uso de testes psicométricos.

Protocolo, estrutura e duração dos treinos

Nenhum dos artigos explicitou uma categoria de protocolo de treino. Todavia, em May et al. (2017) os autores sugerem que estudos futuros utilizem um protocolo do tipo simples-para-complexo em futuros estudos.

Dois artigos utilizaram treinos lineares (Munnelly et al., 2013, 2019). Os demais artigos não relataram a categoria da estrutura de treino. Entretanto, ainda que os autores não tenham relatado explicitamente a categoria, identificamos que quatro artigos utilizaram uma estrutura de um-para-muitos (Dymond et al., 2007, 2008, 2013; Dymond & Whelan, 2010).

Em relação a manipulação experimental da estrutura de treino, Munnelly et al. (2013) avaliaram empiricamente se havia diferenças entre um treino linear em comparação ao treino não-linear, mas não encontraram diferenças estatisticamente significativas entre os grupos expostos aos diferentes formatos.

Quatro artigos não reportaram a quantidade de sessões de treino (Dymond et al., 2007, 2008; Munnelly et al., 2019; Walsh et al., 2014). Os demais experimentos variaram entre uma e duas sessões (Dymond et al., 2013; Dymond & Whelan, 2010; May et al., 2017; Munnelly et al., 2013). Quatro artigos também não reportaram a duração das sessões de treino (Dymond et al., 2007, 2008; Munnelly et al., 2013, 2019). A duração das sessões variou de 15 minutos, no treino mais curto reportado (Walsh et al., 2014), a 120 minutos no treino mais longo reportado (May et al., 2017).

Estímulos utilizados e respostas relacionais

Cinco dos oito artigos utilizaram pseudopalavras no treino (Dymond et al., 2007, 2008, 2013; Dymond & Whelan, 2010; Walsh et al., 2014) e três artigos utilizaram símbolos abstratos (May et al., 2017; Munnelly et al., 2013, 2019). Walsh et al. (2014) variaram estímulos de acordo com participantes, tendo utilizado além de pseudopalavras, imagens variadas e palavras em língua estrangeira.

As dicas contextuais utilizadas variaram entre o uso de símbolos (Dymond et al., 2007, 2008, 2013; Dymond & Whelan, 2010; Munnelly et al., 2013), pseudopalavras (May et al., 2017) ou formas abstratas (Munnelly et al., 2019).

Dois artigos utilizaram estímulos condicionados a funções aversivas em meio aos estímulos neutros (Dymond et al., 2007, 2008). Esses estímulos foram utilizados a fim de treinar o participante a emitir uma resposta de esquiva diante de alguns estímulos previamente neutros (pseudopalavras).

Uso de MET

Somente dois dos oito artigos utilizam o termo "múltiplos exemplares" (May et al., 2017; Walsh et al., 2014). Esses dois também são os únicos que variam o conjunto de estímulos do treino quando o participante não demonstra o DRR, repetindo o ciclo de treino a partir de um novo conjunto de estímulos.

Sete dos oito estudos variaram o conjunto de estímulos em uma fase de treino de dicas contextuais arbitrárias a partir de múltiplas tentativas comparando dimensões não-arbitrárias (Dymond et al., 2007, 2008, 2013; Dymond & Whelan, 2010; May et al., 2017; Munnelly et al., 2013, 2019).

TARPA

Desempenho e critérios de maestria em testes de IC

Três dos quatro experimentos realizaram testes de IC a partir de critérios pré-definidos de maestria (Kishita et al., 2013; Moran et al., 2014, 2015) totalizando 65 participantes na análise.

O TARPA divide-se em percursos com estímulos visuais-visuais, auditivo-visuais e auditivos-auditivos e nem todos os participantes foram expostos aos mesmos percursos no TARPA. O estudo de Kishita et al. (2013) utilizou somente a rota visual-visual e nesse estudo somente um em sete (14%) dos participantes completou o teste de implicação combinatória.

O estudo de Moran et al. (2014) aplicou todas as rotas em uma amostra neurotípica (13 participantes) e uma amostra com diagnóstico de TEA (10 participantes). No grupo neurotípico o desempenho foi de 4 em 13 participantes (30%) na rota visual-visual; 4 em 13 (30%) na rota auditivo-visual 1; 3 em 13 (23%) na rota auditivo-visual 2; e 1 em 13 na rota auditivo-auditivo (7%). Já no grupo diagnosticado com TEA o desempenho foi de 3 em 10 participantes (30%) na rota visual-visual; 2 em 10 (20%) na rota auditivo-visual 1; 2 em 10 (20%) na rota auditivo-visual 2; e 1 em 10 (10%) na rota auditivo-auditivo. O estudo de Moran et al. (2015) utilizou somente a rota auditivo-visual e somente três em 35 participantes (8%) completaram o teste de implicação combinatória. Considerando todos os estudos a média de participantes a completarem testes de IC foi de 19,2% no TARPA. No estudo de Moran et al. (2014) o desempenho em testes de IC foi superior em todos os percursos da amostra neurotípica em relação a amostra com TEA em três das quatro rotas, exceto na rota auditiva-auditiva, onde a porcentagem de participantes a passarem no

teste foram similares. O desempenho foi pior na rota auditiva-auditiva em ambos os grupos, em relação às demais rotas.

Os critérios de maestria variaram entre 83 e 100% entre os estágios dos quatro estudos analisados (Kishita et al., 2013; Moran et al., 2010, 2014, 2015).

Uso de testes psicométricos

Todos os artigos realizaram correlações entre a pontuação dos participantes no TARPA ao final do procedimento e testes psicométricos, sendo os testes: a escala pré-escolar de linguagem (PLS-4; Moran et al., 2014, 2015); a escala Vineland de comportamentos adaptativos (VABS; Moran et al., 2010, 2015); a escala japonesa de maturidade social (Kishita et al., 2013); o teste de inteligência Stanford Binet (SB-5; Moran et al., 2015) e a escala Gilliam de avaliação de comportamentos autistas (Moran et al., 2015). Nenhum dos artigos realizou comparações pré e pós-treino.

Entre a escala de maturidade social e o TARPA (Kishita et al., 2013), foram encontradas correlações de Pearson significativas somente nas sub escalas de comunicação ($r = 0,76$; $p < 0,05$) e socialização ($r = 0,79$; $p < 0,05$). Os demais instrumentos foram analisados a partir de uma análise de correlação de Spearman. Entre PLS-4 e o TARPA foram encontradas correlações de $\rho = 0,96$ ($p < 0,00001$) (Moran et al., 2014) e $\rho = 0,74$ ($p < 0,001$) (Moran et al., 2015).

As correlações entre o VABS e TARPA foram significativas tanto em Moran et al. (2010) ($\rho = 0,97$; $p = 0,05$) quanto em Moran et al. (2015) ($\rho = 0,61$; $p < 0,001$), que também encontrou correlações significativas entre o SB-5 e o TARPA ($\rho = 0,74$; $p < 0,001$).

Protocolo, estrutura e duração dos treinos

Nenhum dos artigos do TARPA reporta categorias de protocolo ou estrutura de treino. Os estudos variaram entre duas e 12 sessões (Kishita et al., 2013; Moran et al., 2010, 2014, 2015) entre o mínimo de aproximadamente 8 minutos (Moran et al., 2015) e máximo de 90 minutos de duração (Moran et al., 2010).

Estímulos utilizados e respostas relacionais

Os estímulos variaram entre sons que emitiam sílabas sem sentido (Moran et al., 2010, 2014, 2015), pseudopalavras (Moran et al., 2010) e formas abstratas (Kishita et al., 2013; Moran et al., 2010, 2014, 2015). Estímulos para dicas contextuais e opções de resposta não foram utilizadas posto que o procedimento incluía somente molduras de igualdade e, portanto, exigia que participantes clicassem em estímulos pertencentes a mesma rede relacional.

Uso do MET

Nenhum dos artigos utiliza o termo múltiplos exemplares ou varia o conjunto de estímulos entre etapas.

T-IRAP

Desempenho e critérios de maestria em testes de IC

Não foram realizados testes de IC no estudo.

Uso de testes psicométricos

Não foram utilizados testes psicométricos no estudo.

Protocolo, estrutura e duração dos treinos

Por não conter testes de IC o estudo não possui categoria de protocolo ou estrutura de treino. As sessões duraram entre 20 e 30 minutos. Não foi reportado o número de sessões por participante (Kilroe et al., 2014).

Estímulos utilizados e respostas relacionais

Como estímulos de treino foram utilizados catálogos de imagens educativas retirados da internet como por exemplo, imagens de objetos (e.g. moedas), animais e representações matemáticas simbolizando "metade" (e.g. 1/2, 50%, ou uma figura geométrica com a metade preenchida). Os autores não reportam a quantidade total de estímulos utilizados. Dois estímulos eram apresentados na tela por vez e como dicas contextuais e opções de resposta foram utilizadas as palavras não-arbitrárias escritas na tela ("igual" ou "diferente" e "mais" ou "menos").

Uso do MET

Os autores citam "múltiplos exemplares" na introdução, mas não no método. Os pesquisadores variaram conjuntos de estímulos no experimento com molduras de igualdade e distinção, mas não variaram conjuntos no experimento com molduras de comparação.

DISCUSSÃO

O objetivo da presente revisão foi comparar variáveis metodológicas e desempenho dos participantes no SMART e em outros treinos automatizados baseados na RFT para desenvolvimento do DRR. Os resultados foram descritos a partir de uma série de parâmetros identificados em cada um dos 27 artigos selecionados: as molduras relacionais treinadas; a amostra; o desempenho e critérios de maestria em testes de IC; o uso de testes psicométricos; o protocolo, estrutura e duração dos treinos; os estímulos utilizados e respostas relacionais e o uso do MET.

Esses parâmetros foram úteis nas comparações de trabalhos intra instrumentos e para apontar lacunas de exploração nos experimentos, mas não são tão eficazes na realização de comparações de eficácia e/ou eficiência entre instrumentos distintos. Isso ocorreu porque os parâmetros divergem entre estudos que utilizam os mesmos treinos e não podem ser avaliados de forma independente. Não se pode, por exemplo, comparar um protocolo a outro se eles possuem amostras muito distintas. Ao mesmo tempo, mesmo que dois trabalhos possuam perfis de amostras similares, eles podem ser diferentes com relação as molduras relacionais utilizadas, critérios de maestria, estímulos utilizados, número de sessões etc. O desempenho dos participantes em testes de IC seria um importante indicador de eficácia de um treino de DRR e é comumente utilizado em estudos com treinos de redes relacionais (Fields et al., 2020). Entretanto, este não pode ser considerado um parâmetro seguro para comparação de eficácia entre treinos quando as demais variáveis entre os estudos não são controladas.

Em relação à amostra, por exemplo, há dificuldade de se estabelecer comparações entre o TARPA e os demais instrumentos. Como demonstrado na Tabela 2, os estudos com esse protocolo utilizam uma amostra muito mais jovem e/ou com diagnóstico de TEA (Kishita et al., 2013; Moran et al., 2010, 2014, 2015). Amostras similares foram utilizadas em dois artigos que utilizaram o REP (Corbett et al., 2017; J. Hayes et al., 2016) mas os participantes não passaram por testes de IC, impedindo a comparação entre o TARPA e o REP. Outros dois estudos utilizaram amostras similares aos estudos com o TARPA, o de Walsh et al. (2014), que utiliza o RCP e o de Kilroe et al. (2014) com o T-IRAP. Esses dois estudos realizaram testes de IC em uma amostra de crianças diagnosticadas com TEA.

Em média somente 19,2% dos participantes atingiram critérios em testes de IC no TARPA (Kishita et al., 2013; Moran et al., 2014, 2015) em comparação com 77% no estudo de Walsh et

al. (2014) com o RCP e 100% no estudo com o T-IRAP (Kilroe et al., 2014). Porém, essas comparações são feitas com algumas ressalvas. Há uma diferença grande na faixa etária dos participantes do TARPA (3 a 12 anos) em relação aos participantes do RCP (5 a 18 anos). Já o estudo de Kilroe et al (2014) com o T-IRAP foi realizado com participantes de 8 a 10 anos mas houve uma diferença no tipo de estímulos utilizados. O T-IRAP se restringiu a estímulos visuais, enquanto os estudos com TARPA utilizaram estímulos auditivos em algumas etapas. O único estudo com o TARPA que utilizou exclusivamente estímulos visuais foi o de Kishita et al. (2013), mas foi realizado com participantes mais jovens, de 2 a 6 anos.

Os dados realçam a lacuna de estudos com participantes mais jovens e neurodiversos dentre os trabalhos aqui encontrados. Somente cinco dos 27 artigos utilizaram amostras em idade pré-escolar (J. Hayes et al., 2016; Kishita et al., 2013; Moran et al., 2014, 2015; Walsh et al., 2014). Ressalta-se que há também uma lacuna de estudos com participantes de meia idade e idosos. Só dois estudos (May et al., 2017; Smeets et al., 2006) incluíram pessoas de 45 a 59 anos e nenhum estudo foi encontrado com populações acima de 59 anos.

As amostras entre alguns estudos do RCP (Dymond et al., 2007, 2008, 2013; Dymond & Whelan, 2010; May et al., 2017; Munnelly et al., 2013, 2019) e REP (Cullinan et al., 2001; Smeets et al., 2004, 2006; Stewart et al., 2004) foram mais semelhantes. Nesses estudos os participantes eram adolescentes e adultos neurotípicos. Ocorreram variações no desempenho em testes de IC a depender das condições experimentais. Em condições experimentais nas quais mais participantes completaram os testes de IC, as porcentagens foram próximas entre RCP e REP. 95% dos participantes no RCP (116 em 122 participantes), 93% (15 de 16 participantes) no REP no formato go/no-go e 100% (17 participantes) no REP no formato com opções de resposta.

Entretanto, essa comparação entre instrumentos em relação aos resultados em testes de IC isoladamente também não é adequada. Além da disparidade no número de participantes, artigos do RCP utilizaram molduras de oposição (Dymond & Whelan, 2010; Dymond et al., 2013) e comparação (Munnelly et al., 2013, 2017) enquanto os artigos do REP limitaram-se a molduras de igualdade e diferença (Cullinan et al., 2001; Smeets et al., 2004, 2006; Stewart et al., 2004). Essa comparação também excluiu o único estudo em toda nossa amostra que empregou molduras relacionais espaciais e portanto não pôde ser comparado aos demais. Neste estudo o desempenho dos participantes em testes de IC no RCP foi muito inferior, um dado preliminar que pode indicar que molduras relacionais espaciais são mais complexas que molduras relacionais de igualdade,

oposição e comparação. Esse dado seria consistente com a literatura que demonstra que diferentes molduras relacionais possuem diferentes níveis de complexidade (O’Hora et al., 2002; Steele & Hayes, 1991). Esse também foi o primeiro estudo a avaliar molduras relacionais espaciais a partir da RFT (May et al., 2017; Dymond & May, 2018) e, por isso, mais estudos precisam avaliar treinos a partir da moldura espacial e as diferenças de complexidade desse tipo de moldura relacional.

As tentativas de comparação acima reforçam como o uso do desempenho em testes de IC foi um parâmetro que possuiu limitações nas comparações entre os instrumentos que encontramos. Fields et al. (2020) também apontam que essa medida considera somente a *eficácia* e não a *eficiência* dos treinos. Não são considerados, por exemplo, quantos blocos de treino foram necessários até que o participante atinja o critério de maestria e ainda que dois treinos diferentes demonstrem que 100% dos participantes completam testes de IC, esse dado não compara quanto treino foi necessário para que o participante conseguisse a completude do teste. Além disso, os critérios de maestria entre diferentes treinos variaram de 75% a 100% de acertos necessários na amostra de artigos analisada.

As comparações entre TARPA, RCP e T-IRAP e entre RCP e REP, foram as únicas possíveis a partir de amostras ligeiramente similares e a partir do desempenho dos participantes em testes de IC. Vale ressaltar que, a partir da leitura dos artigos, descobrimos que o SMART não pôde ser comparado a nenhum treino a partir de testes de IC porque essa medida não foi reportada nos artigos do SMART.

A falta desse dado em treinos SMART pode ser explicada a partir da diferença de objetivo entre os diferentes instrumentos. Os estudos encontrados com TARPA, RCP, REP e T-IRAP objetivaram verificar as variáveis que permitiam a demonstração do DRR. Ou seja, eles focaram na *eficácia* do instrumento e do método em demonstrar o DRR nos participantes e, para isso, a porcentagem de participantes que completam testes de IC foi uma medida útil. Já os estudos com SMART buscaram, como destacado por Colbert et al. (2018), aumentar a *fluência* e possibilidade de generalização do DRR nos participantes a partir de múltiplos exemplares de tentativas e de estímulos.

Essa diferença de objetivos também explica por que estudos com o SMART foram os únicos a utilizarem testes psicométricos como medida pré e pós-treino. Afinal, se o desempenho no DRR está relacionado a testes psicométricos, supõe-se que aumentos no repertório do DRR impliquem em aumentos em medidas de comparação pós-treino (Cassidy et al., 2016). Demais

instrumentos se restringiram a medidas de correlação de desempenho nos treinos com as escalas e desses, dois instrumentos utilizaram escalas comuns entre si. Correlações foram encontradas na escala de linguagem PLS-4 tanto no TARPA (Moran et al., 2014, 2015) quanto no REP (Corbett et al., 2016; J.Hayes et al.,2016). Também foram comuns aos dois treinos as correlações com a escala de inteligência SB-5 (Moran et al., 2015; J. Hayes et al., 2016).

Além dos parâmetros já descritos, buscamos avaliar a categoria dos protocolos, estrutura e duração dos treinos, mas essas variáveis não foram descritas na maior parte dos artigos. A transparência dessas informações metodológicas seria importante tanto para a replicação dos trabalhos e comparação entre instrumentos, impedindo que variáveis confundidoras interfiram nos resultados, quanto para ajudar em propostas de alterações metodológicas nos instrumentos. Como demonstrado, o protocolo de treino foi uma variável independente em dois trabalhos de Smeets et al. (2004, 2006). Os autores demonstraram evidências que os protocolos simples-para-complexo aumentam o número de participantes que completam testes de IC no REP. Já May et al. (2017) não especificam o protocolo de treino, mas sugerem na seção de discussão que próximas pesquisas com o RCP adotem o protocolo simples-para-complexo. Essa exploração não foi realizada em estudos com o TARPA, T-IRAP e SMART. Estudos futuros com esses instrumentos podem comparar se mudanças no protocolo ou estrutura de treino implicarão em mais participantes completando testes de IC ou, no caso do SMART, avançando em mais estágios de treino.

A avaliação do tipo de estímulos utilizados e respostas relacionais exigidas dos participantes demonstrou alguma homogeneidade entre os artigos. Somente um instrumento, o TARPA, destacou-se por utilizar estímulos diferentes dos demais, sendo uma exclusividade do mesmo o uso de estímulos auditivos. Os dados apontaram que o resultado dos participantes em testes de IC foi pior a partir desses estímulos em relação a estímulos visuais (Moran et al., 2010). O uso de estímulos auditivos pode ser avaliado em outros instrumentos a fim de verificar se essa queda de desempenho está relacionada ao tipo dos estímulos ou se é uma característica do TARPA. Além dessa exploração, outros estudos podem verificar se há diferenças no uso de diferentes perfis de estímulos posto que nenhum estudo comparou desempenhos dos participantes comparando símbolos ou formas abstratas a pseudopalavras. Somente o estudo de Stewart et al. (2004) utilizou tanto pseudopalavras quanto formas abstratas no treino. Já o uso de dicas contextuais e opções de resposta quando não utilizados com estímulos não-abstratos (como as palavras "igual" ou "diferente", limitaram-se a símbolos e pseudopalavras. Somente o estudo de Cullinan et al. (2001)

explorou o impacto de diferentes opções de resposta/dicas contextuais no desempenho dos participantes no REP.

Como sugerido por autores do SMART (Cassidy et al., 2016) a variação de diferentes conjuntos de estímulos também podem ser uma demonstração da generalização do DRR. Essa proposição é coerente com a ideia de múltiplos exemplares proposta pela RFT e pontuada na introdução desse trabalho (S.C. Hayes et al., 2001) e portanto, considera-se que uma maior diversidade de estímulos e de opções de resposta seja explorada em trabalhos futuros.

Por último, foi avaliado o uso do MET nos trabalhos. Além do SMART, somente dois artigos do RCP (Walsh et al., 2014; May et al., 2017) utilizaram o MET com um novo conjunto de estímulos durante o treino com estímulos abstratos (e não no pré-treino). Nesses dois trabalhos os participantes que não demonstravam o DRR eram expostos a novos conjuntos de estímulos até atingirem critérios de maestria no IC antes de prosseguirem no restante do protocolo. Nesse sentido, os dois experimentos com o RCP se assimilaram ao SMART metodologicamente. Os resultados demonstraram como alguns participantes que não demonstraram a IC a princípio, conseguiram realizá-lo a partir de treinos com novos conjuntos de estímulos. Todavia, o número de conjuntos de estímulos utilizados no MET foi limitado por restrições de tempo em Walsh et al. (2014) e restrito a um máximo de quatro conjuntos de estímulos em May et al. (2017) o que culminou em alguns participantes serem retirados dos procedimentos antes de demonstrarem o DRR. Ainda assim, esses estudos com o RCP demonstraram o potencial do MET no desenvolvimento do DRR em indivíduos que não exibem repertório. Não é possível saber se, por exemplo, o baixo índice de participantes que completaram testes de IC no TARPA poderia ter sido melhorado a partir de um treino com MET. Essa hipótese precisa ser explorada nos próximos trabalhos com este e com demais instrumentos.

As diferenças nos estudos discutidas aqui não impedem uma comparação entre alguns instrumentos no futuro. Os resultados dessa revisão destacam que a partir da amostra de estudos selecionada e a partir dos parâmetros analisados, a análise comparativa entre as variáveis metodológicas apresentou dificuldades. Não houve interseções suficientes entre os parâmetros que destacamos nos estudos para que eles fossem comparados entre si em relação ao desempenho dos participantes em testes de IC, mas isso não implica que essas comparações não possam ser realizadas no futuro a partir de ajustes metodológicos. Por exemplo, seria possível uma comparação de grupos entre TARPA, RCP, REP e T-IRAP se fossem utilizados os mesmos

estímulos, molduras, amostras similares e critérios de maestria. Já com relação ao SMART consideramos que essa comparação é mais difícil devido às diferenças de objetivos encontradas entre esse e outros instrumentos.

Futuros trabalhos também podem tentar propor alternativas com objetivos similares ao SMART, que foquem no treino para a generalização e aumento de habilidades relacionadas ao DRR e não somente na eficácia de formação de redes relacionais pelos participantes. Para isso, uma via possível é que estudos com demais instrumentos utilizem um formato de treino mais similar ao SMART, utilizando o MET com múltiplos conjuntos de estímulos e comparando medidas psicométricas no pré e pós-treino. Além disso, os resultados apontaram que outras variáveis podem ser investigadas tanto no SMART como nos outros instrumentos através de aplicações dos treinos em amostras mais diversas, o uso de diferentes protocolos e estruturas de treino e a variação do tipo dos estímulos utilizados tanto no treino quanto nas opções de resposta.

Finalmente, há limitações no presente trabalho que precisam ser discutidas. Os parâmetros escolhidos podem não ter sido suficientes para uma comparação adequada entre os diferentes instrumentos. Futuros trabalhos podem propor novos parâmetros para comparação ao invés dos escolhidos pelos autores. Apesar de não ter sido avaliada no presente trabalho, a transformação de função é uma propriedade importante do DRR e já foi avaliada em alguns estudos (Dymond et al. 2007, 2008; Moran et al., 2014; 2015).

Em nosso trabalho, nenhum parâmetro de avaliação que envolvesse especificamente a qualidade de tecnologias de automação foi utilizado, como o uso de intervenções remotas ou comparações entre diferentes aparatos (e.g. comparar um treino com tela sensível ao toque ao invés de uso de *mouse* e teclado). Ainda assim, esperava-se que, dada a busca por protocolos automatizados, mais fosse dito a respeito das peculiaridades dessas intervenções. Contudo, somente um dos trabalhos comparou uma intervenção automatizada a uma intervenção não automatizada (Kilroe et al., 2014). Nesse estudo os participantes demonstraram respostas mais rápidas e acuradas em tentativas de implicação mútua e tentativas de IC na intervenção mediada por software em relação ao procedimento de mesa. Esse resultado, ainda que preliminar, mostra o potencial de treinos automatizados em relação aos procedimentos não automatizados.

A última limitação do trabalho relaciona-se ao método utilizado para seleção dos artigos. A maior parte dos artigos (26 dos 27 artigos, ver Figura 2) foram encontrados através do método *snowballing* (Wohlin, 2014). Somente um dos artigos da amostra foi exclusivo da busca através

de bases de dados. Por um lado, isso indica o potencial do método *snowballing* para identificar artigos que não seriam encontrados facilmente através de uma busca nas bases de dados, por outro, a seleção desses artigos foi realizada individualmente pelo pesquisador e, portanto, há possibilidades de erros e vieses na seleção dos artigos, além de dificultar a replicação por outros pesquisadores. A busca sistemática em bases de dados a partir da seleção de três autores distintos foi uma tentativa de reduzir possíveis vieses, mas essa seleção não foi realizada de forma cega e, apesar de terem sido utilizadas várias bases, retornou poucos resultados em comparação com o *snowballing*.

CONCLUSÃO

Em síntese, esse foi o primeiro trabalho encontrado que buscou identificar e comparar características metodológicas e resultados dos protocolos automatizados baseados na RFT encontrados entre 2001 e março de 2021. Em geral os resultados sugeriram uma heterogeneidade que dificultou a comparação entre os artigos, mas foram destacadas uma série de variáveis independentes que podem ser manipuladas em novos estudos. Para o avanço no conhecimento sobre o impacto dos protocolos automatizados de treino baseados na RFT há necessidade de que os estudos sejam realizados a partir de um diálogo metodológico entre as diferentes pesquisas.

REFERÊNCIAS

- Amd, M., & Roche, B. (2018). Assessing the Effects of a Relational Training Intervention on Fluid Intelligence Among a Sample of Socially Disadvantaged Children in Bangladesh. *The Psychological Record*, 68(2), 141–149. <https://doi.org/10.1007/s40732-018-0273-4>
- Arntzen, E. (2012). Training and Testing Parameters in Formation of Stimulus Equivalence: Methodological Issues. *European Journal of Behavior Analysis*, 13(1), 123–135. <https://doi.org/10.1080/15021149.2012.11434412>
- Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences*, 47, 222–235. <https://doi.org/10.1016/j.lindif.2016.03.001>
- Cassidy, S., Roche, B., & Hayes, S. C. (2011). A Relational Frame Training Intervention to Raise Intelligence Quotients: A Pilot Study. *The Psychological Record*, 61(2), 173–198. <https://doi.org/10.1007/BF03395755>
- Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART Training Really Increase Intelligence? A Replication Study. *Journal of Behavioral Education*, 27(4), 509–531. <https://doi.org/10.1007/s10864-018-9302-2>
- Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis. *Qualitative Health Research*, 22(10), 1435–1443. <https://doi.org/10.1177/1049732312452938>
- Corbett, O., Hayes, J., Stewart, I., & McElwee, J. (2017). Assessing and training children with autism spectrum disorder using the relational evaluation procedure (REP). *Journal of Contextual Behavioral Science*, 6(2), 202–207. <https://doi.org/10.1016/j.jcbs.2017.02.007>
- Cullinan, V. A., Barnes-Holmes, D., & Smeets, P. M. (2001). A PRECURSOR TO THE RELATIONAL EVALUATION PROCEDURE: SEARCHING FOR THE CONTEXTUAL CUES THAT CONTROL EQUIVALENCE RESPONDING. *Journal of the Experimental Analysis of Behavior*, 76(3), 339–349. <https://doi.org/10.1901/jeab.2001.76-339>
- Dymond, S., & May, R. (2018). Quantifying the empirical growth of relational frame theory research: A cautionary note. *The Psychological Record*, 68(2), 255–260.
- Dymond, S., Ng, T. C., & Whelan, R. (2013). Establishing Arbitrarily Applicable Relations of Same and Opposite With the Relational Completion Procedure: Selection-Based Feedback. *The Psychological Record*, 63(1), 111–130. <https://doi.org/10.11133/j.tpr.2013.63.1.009>
- Fields, L., Arntzen, E., & Doran, E. (2020). Yield as an Essential Measure of Equivalence Class Formation, Other Measures, and New Determinants. *The Psychological Record*, 70(2), 175–186. <https://doi.org/10.1007/s40732-020-00377-3>
- Fields, L., Hobbie-Reeve, S. A., Adams, B. J., & Reeve, K. F. (1999). Effects of Training Directionality and Class Size on Equivalence Class Formation by Adults. *The Psychological Record*, 49(4), 703–723. <https://doi.org/10.1007/BF03395336>
- Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2018). Exploring the Single-Trial-Type-Dominance-Effect in the IRAP: Developing a Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) Model. *The Psychological Record*, 68(1), 11–25. <https://doi.org/10.1007/s40732-017-0262-z>
- Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *British Journal of Educational Psychology*, 86(3), 397–411. <https://doi.org/10.1111/bjep.12114>

- Hayes, J., Stewart, I., & McElwee, J. (2016). Assessing and Training Young Children in Same and Different Relations Using the Relational Evaluation Procedure (REP). *The Psychological Record*, 66(4), 547–561. <https://doi.org/10.1007/s40732-016-0191-2>
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. Springer US. <https://doi.org/10.1007/b108413>
- Holland, J. G. (1960). TEACHING MACHINES: AN APPLICATION OF PRINCIPLES FROM THE LABORATORY ¹. *Journal of the Experimental Analysis of Behavior*, 3(4), 275–287. <https://doi.org/10.1901/jeab.1960.3-275>
- Holth, P. (2017). Multiple Exemplar Training: Some Strengths and Limitations. *The Behavior Analyst*, 40(1), 225–241. <https://doi.org/10.1007/s40614-017-0083-z>
- Kilroe, H., Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2014). Using the T-IRAP interactive computer program and applied behavior analysis to teach relational responding in children with autism. *Behavioral Development Bulletin*, 19(2), 60–80. <https://doi.org/10.1037/h0100578>
- Kishita, N., Ohtsuki, T., & Stewart, I. (2013). The Training and Assessment of Relational Precursors and Abilities (TARPA): A follow-up study with typically developing children. *Journal of Contextual Behavioral Science*, 2(1–2), 15–21. <https://doi.org/10.1016/j.jcbs.2013.01.001>
- LaFrance, D. L., & Tarbox, J. (2019). The importance of multiple exemplar instruction in the establishment of novel verbal behavior. *Journal of Applied Behavior Analysis*, jaba.611. <https://doi.org/10.1002/jaba.611>
- Magiati, I., & Howlin, P. A. (2001). Monitoring the progress of preschool children with autism enrolled in early intervention programmes: Problems in cognitive assessment. *Autism*, 5, 399–406.
- May, R. J., Stewart, I., Baez, L., Freegard, G., & Dymond, S. (2017). Arbitrarily applicable spatial relational responding: SPATIAL RELATIONS. *Journal of the Experimental Analysis of Behavior*, 107(2), 234–257. <https://doi.org/10.1002/jeab.250>
- McLoughlin, S., Tyndall, I., & Pereira, A. (2018). Piloting a brief relational operant training program: Analyses of response latencies and intelligence test performance. *European Journal of Behavior Analysis*, 19(2), 228–246. <https://doi.org/10.1080/15021149.2018.1507087>
- McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Relational Operant Skills Training Increases Standardized Matrices Scores in Adolescents: A Stratified Active-Controlled Trial. *Journal of Behavioral Education*. <https://doi.org/10.1007/s10864-020-09399-x>
- McLoughlin, S., Tyndall, I., Pereira, A., & Mulhern, T. (2021). Non-verbal IQ Gains from Relational Operant Training Explain Variance in Educational Attainment: An Active-Controlled Feasibility Study. *Journal of Cognitive Enhancement*, 5(1), 35–50. <https://doi.org/10.1007/s41465-020-00187-z>
- Ming, S., Stewart, I., McElwee, J., & Bynum, K. (2015). Contextual control over derived relational responding in a teenager with autism. *Research in Autism Spectrum Disorders*, 19, 7–17. <https://doi.org/10.1016/j.rasd.2015.03.003>
- Moran, L., Stewart, I., McElwee, J., & Ming, S. (2010). Brief Report: The Training and Assessment of Relational Precursors and Abilities (TARPA): A Preliminary Analysis. *Journal of Autism and Developmental Disorders*, 40(9), 1149–1153. <https://doi.org/10.1007/s10803-010-0968-0>
- Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational Ability and Language Performance in Children With Autism Spectrum Disorders and Typically Developing Children: A Further Test of the TARPA Protocol. *The Psychological Record*, 64(2), 233–251. <https://doi.org/10.1007/s40732-014-0032-0>

- Moran, L., Walsh, L., Stewart, I., McElwee, J., & Ming, S. (2015a). Correlating derived relational responding with linguistic and cognitive ability in children with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *19*, 32–43. <https://doi.org/10.1016/j.rasd.2014.12.015>
- Munnely, A., Freegard, G., & Dymond, S. (2013). Constructing Relational Sentences: Establishing Arbitrarily Applicable Comparative Relations with the Relational Completion Procedure. *The Psychological Record*, *63*(4), 751–768. <https://doi.org/10.11133/j.tpr.2013.63.4.004>
- O’hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived Relational Responding and Performance on Verbal Subtests of the WAIS-III. *The Psychological Record*, *55*(1), 155–175. <https://doi.org/10.1007/BF03395504>
- O’Hora, D., Peláez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal Relations and Intelligence: Correlating Relational Performance With Performance on the Wais-III. *The Psychological Record*, *58*(4), 569–584. <https://doi.org/10.1007/BF03395638>
- O’Toole, C., & Barnes-Holmes, D. (2009). Three Chronometric Indices of Relational Responding as Predictors of Performance on a Brief Intelligence Test: The Importance of Relational Flexibility. *The Psychological Record*, *59*(1), 119–132. <https://doi.org/10.1007/BF03395652>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>
- Page, M. J., Moher, D., Bossuyt, P., Boutron, I., Hoffmann, T., mulrow, cindy, Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. (2020). *PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews* [Preprint]. MetaArXiv. <https://osf.io/gwdhk>
- Reed, P., Osborne, L., & Corness, M. (2005). The effectiveness of early intervention programmes for autistic spectrum disorders. A Report for the South East Regional Special Educational Needs Partnership. Research Partners: Bexley, Brighton & Hove, East Sussex, Kent, Midway, Surrey, West Sussex.
- Skinner, B. F. (1958). Teaching Machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science*, *128*(3330), 969–977. <https://doi.org/10.1126/science.128.3330.969>
- Smeets, P. M., Barnes-Holmes, D., & Striefel, S. (2006). Establishing and Reversing Equivalence Relations with a Precursor to the Relational Evaluation Procedure. *The Psychological Record*, *56*(2), 267–286. <https://doi.org/10.1007/BF03395550>
- Smeets, P. M., Wijngaarden, M. van, Barnes-Holmes, D., & Cullinan, V. (2004). Assessing stimulus equivalence with a precursor to the relational evaluation procedure. *Behavioural Processes*, *65*(3), 241–251. <https://doi.org/10.1016/j.beproc.2003.10.004>
- Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the Relational Evaluation Procedure. *The Psychological Record*, *54*(4), 531–552. <https://doi.org/10.1007/BF03395491>
- Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational frame theory, mathematical, and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology & Psychological Therapy*, *16*(2), 141–155.
- Walsh, S., Horgan, J., May, R. J., Dymond, S., & Whelan, R. (2014). Facilitating relational framing in children and individuals with developmental delay using the relational completion procedure: FACILITATING RELATIONAL FRAMING. *Journal of the Experimental Analysis of Behavior*, *101*(1), 51–60. <https://doi.org/10.1002/jeab.66>

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, 1–10.
<https://doi.org/10.1145/2601248.2601268>

Estudo 2

Há como aumentar os escores dos participantes no SMART? Uma investigação do uso de múltiplos estímulos e parâmetros de tentativas de treino

O treino chamado SMART foi desenvolvido a partir de um trabalho conduzido por Cassidy et al. em 2011. Os participantes expostos a esse treino de responder relacional, baseado na Teoria das Molduras Relacionais (S. C. Hayes et al., 2001), têm demonstrado aumento de escores em testes psicométricos e de habilidades acadêmicas em diversos estudos (Amd & Roche, 2018; Cassidy et al., 2016; J. Hayes & Stewart, 2016; McLoughlin et al., 2020, 2021; Thirus et al., 2016). McLoughlin e colaboradores (2020) apontaram que participantes treinados no SMART exibiram mais ganhos do que os relatados no treino mais popular de memória de trabalho.

O SMART baseia-se na premissa de que, a partir de um treino com múltiplos exemplares (MET), o participante melhorará seu repertório no responder relacional derivado (DRR; S.C. Hayes et al., 2001) e, portanto, melhorará habilidades cognitivas (Cassidy et al., 2016). Essa teoria parece ser consistente com estudos anteriores que demonstram correlações entre o DRR e medidas em escalas psicométricas (O’Hora et al., 2005, 2008; O’Toole & Barnes-Holmes, 2009).

Apesar das evidências favoráveis ao uso do SMART, há relatos que apontam participantes que não conseguem completar as tentativas do treino (Amd & Roche, 2018; J. Hayes & Stewart, 2016). O estudo de Amd & Roche (2018) fornece alguns dados sobre a quantidade de avanço dos participantes e demonstra que, mesmo com no mínimo vinte sessões de treino, 10 de 35 participantes não conseguiram ultrapassar sequer o primeiro estágio de treino. Entretanto, não foram discutidos argumentos que apontem as razões dessa dificuldade no avanço no treino, mas há outras dificuldades envolvidas no método utilizado que podem explicar esse baixo desempenho (como o uso de inglês com participantes que não eram fluentes na língua). Porém, esse estudo foi o único encontrado que detalhou quantos participantes não avançaram em cada estágio.

Até então, as únicas evidências que explicam dificuldades dos participantes com relação ao SMART são relacionadas a aspectos mensurados por testes de personalidade (McLoughlin et al., 2020, 2021). Nenhuma pesquisa investigou o porquê de alguns participantes não conseguirem avançar em alguns estágios a partir de questionamentos sobre características do próprio protocolo de treino. O objetivo do presente estudo envolve essa investigação. Apontaremos primeiramente

um breve resumo sobre o funcionamento do SMART e discutiremos duas possibilidades metodológicas de mudança no instrumento: alterações no uso do MET e na ordem de apresentação das tentativas, a serem investigadas experimentalmente com uma adaptação de jogo, baseada no SMART.

O SMART é composto de tentativas que variam de acordo com os parâmetros de 55 estágios (para uma descrição completa dos parâmetros e estágios, ver Cassidy et al. [2016]). Cada tentativa é composta de um número pré-determinado de premissas com pseudopalavras (e.g. JOL, REB, SIG) e uma pergunta sobre elas. Esses parâmetros podem ser explicados a partir dos exemplos dos três quadros da Tabela 1, nela são demonstradas tentativas dos estágios 27, 28 e 29.

Tabela 1

Três exemplos de tentativas utilizadas nos estágios 27, 28 e 29 do SMART

Estágio 27	Estágio 28	Estágio 29
A é igual a B.	A é oposto a B.	A é oposto a B.
B é oposto a C.	B é igual a C.	B é oposto a C.
C é oposto a D.	C é igual a D.	C é igual a D.
A é oposto a D?	A é igual a D?	D é igual a B?

Nota. As letras são substituídas por pseudopalavras no treino (e.g. *XOJ*, *HEQ* etc). Da esquerda para a direita, uma tentativa baseada no estágio 27, uma no estágio 28 e uma no estágio 29 do SMART. As premissas são as mesmas entre o mesmo estágio, mas diferentes parâmetros podem ser utilizados para as questões (ver Tabela 2).

O conjunto de pseudopalavras utilizado nunca se repete entre uma tentativa e outra. Os parâmetros definidores das tentativas do SMART e que utilizamos como base para nossa proposta metodológica são a distância nodal entre os estímulos na questão, a direção da questão, o tipo de relação presente na questão e as relações presentes nas premissas

Na Tabela 1, por exemplo, no primeiro quadro, a pergunta envolve uma distância nodal entre os estímulos de dois nodos no primeiro quadro. Nele, é perguntado se “A é igual a D”. A distância entre A e D é de dois estímulos, B e C, por isso a tentativa utiliza uma distância nodal de dois nodos. Já no terceiro quadro da Tabela 1 a pergunta envolve a distância de um único nodo (é perguntado sobre B e D, e só C está entre B e D). A direção da questão é linear nos dois primeiros quadros e invertida no terceiro quadro. Ou seja, nos dois primeiros quadros as questões seguem a

ordem de apresentação dos estímulos, enquanto no terceiro quadro a questão é invertida (pergunta-se “D é igual a B” ao invés de perguntar se “B é igual a D”). Já o tipo de relação da questão varia entre questões de igualdade e questões de oposição. No primeiro quadro da Tabela 1, a relação da questão é de igualdade (“A é igual a D?”), enquanto no segundo quadro a relação da questão já é de oposição (“A é oposto a D?”).

No SMART cada estágio possui diferentes premissas. O estágio 27 utiliza, nessa ordem, uma premissa de igualdade (“A é igual a B”) e duas de oposição (“B é oposto a C.”). Já o estágio 28 utiliza uma premissa de oposição e duas de igualdade, respectivamente. O estágio 29 utiliza duas premissas de oposição e uma de igualdade, nessa ordem.

Devido à variação dos estímulos abstratos utilizados no treino do SMART, a ênfase das tentativas é colocada sobre o treino das relações entre os estímulos e não nos estímulos em si (Cassidy et al., 2016, Colbert et al., 2018). No SMART o MET é composto então de uma variação de conjuntos de estímulos, as molduras relacionais utilizadas e os graus de complexidade definidos pelos parâmetros. Porém, ainda que esse seja o procedimento empregado por estudos com o SMART, um procedimento padrão do MET não é bem definido na literatura, como apontado por alguns autores (Holth, 2017; LaFrance & Tarbox, 2019).

Como exemplo da falta de consenso quanto à definição há o estudo de Walsh et al. (2014), que também afirmou ter utilizado múltiplos exemplares, mas utilizou um procedimento diferente do SMART. O estudo foi realizado de tal forma que os participantes precisavam formar uma série de relações entre o mesmo conjunto de estímulos antes de iniciarem em um conjunto novo. Nesse estudo os participantes eram treinados em uma série de relações de base ($A \rightarrow B$ e $A \rightarrow C$) e depois eram testados em tentativas de implicação combinatória ($B \rightarrow C$ e $C \rightarrow B$). Caso falhassem no teste, os participantes eram treinados, com o mesmo conjunto, em tentativas de implicação combinatória (com reforço). Quando os participassem atingiam o critério de 90% de acertos por bloco de treino eles eram, então, expostos a um treino de relações de base com um novo conjunto de estímulos e eram testados novamente com relação a implicação combinatória. Em caso de falha, o processo se repetia com novos conjuntos de estímulos.

Ou seja, notamos que há uma diferença metodológica no uso de múltiplos exemplares como empregado por Walsh et al. (2014) em relação ao uso de múltiplos exemplares no SMART. Enquanto no SMART os autores variam o conjunto de estímulos a cada tentativa, no estudo de Walsh et al. (2014) os experimentadores continuam treinando relações com o mesmo conjunto de

estímulos até serem atingidos critérios de maestria, antes dos participantes serem testados em um conjunto de estímulos novos.

Não encontramos trabalhos que comparem, empiricamente, a diferença entre esses dois perfis de treino de múltiplos exemplares nem mesmo dentre os estudos que debatem o conceito de MET, incluindo um artigo de revisão sobre o tema (Holth, 2017; LaFrance & Tarbox, 2019). Isso coloca questões procedimentais a serem investigadas, relacionadas ao fato de que talvez o procedimento de MET como realizado a partir do SMART não seja o que permita que o maior número de participantes conclua o treino, como relatado em Amd e Roche (2018) quando alguns participantes não conseguiram avançar em alguns estágios.

Também é possível que primeiramente um participante precise ser reforçado em respostas de implicação combinatória com um único conjunto de estímulos antes que possa ser reforçado com diversos conjuntos, ou seja, a ordem dos diferentes métodos de treino pode importar. Considerando isso, o primeiro objetivo do nosso trabalho foi verificar se a mudança de estímulos por tentativa impacta no escore dos participantes (número de acertos).

A nossa primeira hipótese foi que o uso do MET mantendo o mesmo conjunto de estímulos ao longo das tentativas estará correlacionado a um maior escore dos participantes no estágio final, em comparação com participantes cujo MET varia o conjunto de estímulos a cada tentativa. Essa hipótese é embasada na literatura que demonstra melhorias no DRR mesmo em participantes com atrasos no desenvolvimento a partir do uso de MET treinando o conjunto de estímulos antes de treinar um conjunto novo (Walsh et al., 2014). O estudo de Amd & Roche (2018), realizado com uma amostra sem diagnóstico, demonstrou que 28% dos participantes não conseguiram passar sequer do primeiro estágio do SMART. Neste estudo, a amostra era composta de crianças em situação de vulnerabilidade socioeconômica, aspecto que pode ter impactado no número de participantes que avançaram no treino. Para buscar investigar o possível impacto dessa variável em diferenças no escore dos participantes, no nosso estudo utilizamos um questionário socioeconômico, o Critério Brasil (Associação Brasileira de Empresas de Pesquisa, 2021) segundo o qual maior o escore maior a estimativa de renda familiar. Foi investigada a correlação entre o escore dos participantes no treino e o escore no questionário.

Além disso, os autores afirmam que os parâmetros do SMART entre um estágio e outro variam de acordo com a ordem de complexidade, de forma progressiva (Cassidy et al., 2016; McLoughlin et al., 2021). Contudo, não são apontadas evidências empíricas que embasam a ordem

dos treinos apresentados. Ou seja, é possível que estágios iniciais do treino sejam mais difíceis que estágios mais avançados. Como consequência teríamos que os participantes não avancem certos estágios pode estar relacionado a falta de repertório deles, dada a ordem não embasada de apresentação das tentativas de treino.

Portanto, também buscamos investigar se a ordem de apresentação dos estágios do SMART é coerente com a proposta de uma progressão gradual de dificuldade. Para testarmos a hipótese de que o treino não ocorre segundo essa progressão, comparamos o desempenho dos participantes em tentativas baseadas em três estágios do SMART, os estágios 27, 28 e 29. A suposição é que serão encontrados escores superiores no estágio 28 em relação aos estágios 27 e 29. O estágio 28 (ver Tabela 1) possui duas premissas de igualdade e uma de oposição em comparação com os estágios 27 e 29, que contém duas premissas de oposição e uma de igualdade. Essa hipótese tem como base a literatura que aponta que relações de igualdades são mais básicas que relações de oposição (O’Hora et al., 2002; Steele & Hayes, 1991). Toda a sequência de tentativas segue exatamente os mesmos parâmetros desses três estágios do SMART como pontuado em Cassidy et al. (2016). Todas as possibilidades de questões dos estágios 27, 28 e 29 do SMART são apresentadas na Tabela 2, três exemplos de cada um dos estágios foram apontados na Tabela 1.

Tabela 2

Todas as doze possibilidades de questão dos três estágios do SMART (27,28 e 29)

A é igual a C?	C é igual a A?
B é igual a D?	D é igual a B?
A é igual a D?	D é igual a A?
A é oposto a C?	C é oposto a A?
B é oposto a D?	D é oposto a B?
A é oposto a D?	D é oposto a A?

Nota. As letras são substituídas por pseudopalavras nas tentativas. Os parâmetros das questões variam de acordo com a distância nodal da questão (1 ou 2), direção da questão (linear ou invertida) e relação presente na questão (*igual* ou *oposto*).

No presente estudo o desempenho foi medido a partir do número de acertos por estágio (escore). Escolhemos os três estágios (27, 28 e 29) para a tentativa pois eles utilizam parâmetros similares para as questões apresentadas aos participantes. Esses parâmetros são a distância nodal de 1 ou 2 nodos, relações de igualdade ou oposição na questão e direção da questão igual ou invertida (ver Tabela 2). A única diferença entre os estágios é o número de premissas de igualdade e oposição e a ordem de apresentação dessas premissas (ver Tabela 1). A escolha desses três estágios também se apoia no fato de que no estudo de Amd & Roche (2018), ao longo de trinta sessões de treino, o participante que mais avançou no treino chegou ao estágio 30. Consideramos pertinente, então, que utilizássemos tentativas baseadas nos estágios cujo nível de dificuldade fosse elevado o suficiente para demonstrar algum tamanho de efeito ao comparar o desempenho dos participantes em cada um deles, mas não tão elevados a ponto de impedir que os participantes sequer conseguissem responder.

Outro aspecto com relação ao funcionamento do SMART é que cada um dos seus estágios apresenta tentativas aleatoriamente, a partir dos parâmetros de cada estágio (distância nodal na questão, direção da questão e relação da questão). Isso quer dizer que no treino tradicional dois participantes podem ter experiências diferentes ao realizar as tentativas, ainda que estejam exatamente no mesmo estágio. Essa diferença entre aspectos do SMART é problemática tendo em vista que é esperado que diferentes parâmetros aumentem ou diminuam o grau de dificuldade da tentativa e, portanto, os participantes podem passar por uma seleção de atividades mais difíceis e outros por mais fáceis, de modo aleatório. Assim, todos os participantes do presente estudo realizaram todas as possibilidades de tentativas dentro do mesmo estágio e na mesma ordem. Realizamos análises para comparação da probabilidade de acerto de cada parâmetro por participante. Encontrando dados significativos quanto a estes aspectos, poderia ser sugerida uma ordem de tentativas para procedimentos futuros baseados no SMART, fundamentada no aumento gradual de dificuldade.

Conforme descrito, não identificamos na literatura respostas para algumas questões metodológicas levantadas sobre o SMART. Essas questões envolvem possíveis explicações para o fato de alguns participantes não conseguirem avançar em alguns estágios no protocolo de treino. Para tanto, o objetivo do presente estudo foi verificar se o uso de estímulos repetidos ou variados por estágio resulta em maior escore dos participantes ao final do treino e verificar se a ordem de

estágios e de tentativas do SMART aumentam gradualmente o nível de dificuldade. Adaptamos três estágios do SMART para um jogo computadorizado no qual o escore dos participantes foi avaliado por estágio em quatro condições experimentais. Avaliamos as chances de acerto em tentativas de acordo com os diferentes parâmetros em um mesmo estágio, utilizado como linha de base nas quatro condições.

MÉTODO

Participantes

Selecionamos 43 participantes para o estudo, crianças entre 9 e 12 anos de ambos os sexos sem diagnósticos prévios de condições psiquiátricas. Excluimos dois participantes da amostra devido a um erro no software durante a aplicação. Excluimos outro participante pois este solicitou interrupção no meio do procedimento. Ao final da pesquisa analisamos os dados de 40 participantes.

Os responsáveis pelos participantes foram convidados por meio de divulgação em redes sociais dos pesquisadores e de colegas do laboratório de pesquisa, por meio de contato pessoal e por convite a uma escola municipal de Belo Horizonte. Este último foi autorizado pela Secretaria Municipal de Educação e coordenação pedagógica da escola, seguindo os trâmites exigidos pelo Conselho de ética em pesquisa da UFMG (CEP/UFMG). Os responsáveis pelas crianças preencheram o Termo de Consentimento Livre e Esclarecido (TCLE, Anexo 1) antes do início da coleta de dados. Antes do início dos procedimentos, o experimento era explicado para os participantes e os pesquisadores liam e preenchiam junto a eles, um Termo de Assentimento Livre e Esclarecido (TALE, Anexo 2). Era informado aos participantes que eles poderiam parar ou interromper o procedimento a qualquer momento, caso assim desejassem.

Instrumentos

Critério Brasil

O Critério Brasil é um instrumento de avaliação socioeconômica (Associação Brasileira de Empresas de Pesquisa, 2021) composto de perguntas com relação aos itens de conforto (e.g. quantidade de automóveis que a família possui), grau de instrução do chefe da família e características sobre a moradia (e.g. se a rua é pavimentada e se a água é proveniente de rede geral de distribuição). A partir desses itens, é pontuado um escore que provê informações sobre os estratos socioeconômicos dos participantes. Quanto maior a pontuação, maior a estimativa da renda média domiciliar. Os responsáveis pelos participantes preencheram o Critério Brasil em um formulário online, previamente a aplicação do procedimento experimental.

Jogo adaptado a partir do SMART

Programamos nosso jogo baseado no SMART com o auxílio de um programador externo, a partir da plataforma de desenvolvimento de jogos Unity (www.unity.com). Os participantes

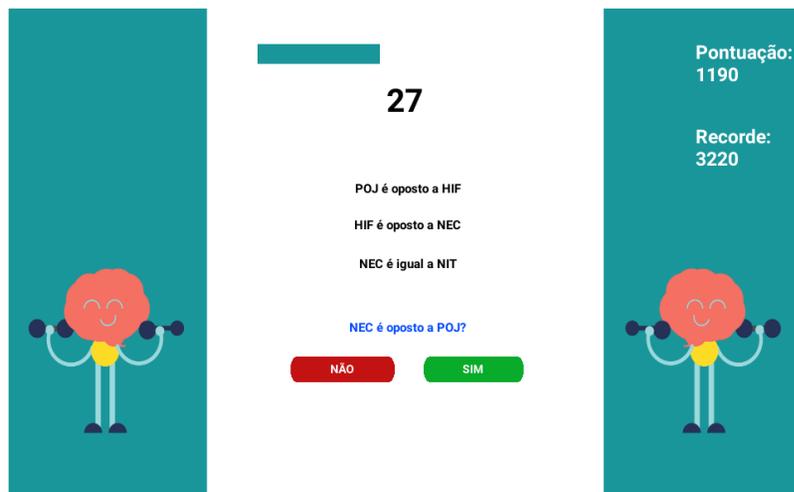
tinham acesso remoto ao jogo através de um site e um ID fornecido pelo experimentador. Criamos as imagens do treino através da plataforma Adobe Illustrator.

Para definição dos parâmetros do treino utilizamos os dados fornecidos por Cassidy et al. (2016). O jogo continha ao todo 48 tentativas, divididas em quatro estágios de 12 tentativas cada. Cada tentativa, assim como no SMART, era composta por três premissas afirmativas e uma pergunta para a qual o participante deveria responder “sim” ou “não” (ver Figura 1). O primeiro estágio foi sem reforço, para estabelecimento de uma linha de base individual, e os três seguintes com reforço. O participante tinha até 30 segundos para responder cada tentativa e, portanto, cada estágio levou no máximo 6 minutos para sua completude. O tempo total do procedimento limitou-se ao máximo de 24 minutos. Para seleção das tentativas utilizamos os parâmetros dos três estágios dentre os 55 do SMART tradicional, os estágios 27, 28 e 29 (ver Tabelas 1 e 2). O primeiro estágio, de linha de base, utilizava as mesmas tentativas do estágio 27 do SMART, sem reforço. O segundo estágio, utilizava as mesmas tentativas do estágio 27, porém com reforço. O terceiro e quarto estágios utilizavam, em sequência, as tentativas dos estágios 28 e 29.

Os parâmetros que nortearam todas as tentativas, sem exceção, foi a composição baseada em três premissas, com questões de igualdade ou oposição dispostas de forma linear ou invertida e com perguntas de distância nodal 1 ou 2. O que variou entre os estágios é a sequência das premissas e a relação entre elas. Os estágios 1 e 2 do nosso jogo, assim como o estágio 27 do SMART, contém a primeira premissa de igualdade, enquanto a segunda e a terceira são de oposição. O estágio 3 contém a primeira premissa de oposição, enquanto a segunda e a terceira são de igualdade. O estágio 4 contém a primeira e segunda premissas de oposição, enquanto a terceira é de igualdade. A Figura 1 demonstra a tela vista pelo participante, com uma tentativa do estágio final do jogo.

Figura 1

Tela de exemplo do estágio final do jogo utilizado na coleta de dados, adaptado a partir do SMART.



Na Figura 1 está exemplificada uma tentativa das 12 possibilidades do último estágio, a fim de explicitarmos os parâmetros utilizados, a resposta correta é “não”. Como se trata do último estágio, a primeira e segunda premissas são de oposição, enquanto a última é de igualdade. A pergunta utiliza a relação de oposição e a direção da pergunta está invertida em relação a direção das premissas. A distância nodal é 1 (há somente um estímulo, “HIJ”, intermediando a relação entre “NEC” e “POJ”). No topo da tela era exibido para os participantes uma barra de progresso, que avançava a cada tentativa e demonstrava para o jogador quanto faltava para acabar o estágio. No canto superior direito era exibida uma pontuação e abaixo dela quanto o jogador pontuou no último estágio. Em caso de acerto o jogador também ganhava um bônus de pontuação por tempo, informação que era dada a ele previamente ao início do jogo.

Procedimentos

Pré treino

Todo o procedimento era acompanhado por uma videochamada de um dos experimentadores, utilizando as plataformas Zoom ou Google Meet. Além do pesquisador principal, os demais experimentadores eram duas estudantes de iniciação científica previamente treinadas. No início da videochamada os participantes eram requisitados a compartilharem sua tela e o experimentador os orientava durante todo o experimento.

Antes do início do treino os participantes passavam por um tutorial explicativo feito verbalmente pelo experimentador e por um pré-treino que demonstrava à criança como operar o jogo e testava sua habilidade na tentativa que envolvia relacionar estímulos não-abstratos. O tutorial era composto de uma apresentação de slides onde era apresentado ao participante três premissas: "Sol é igual a dia", "lua é igual a noite" e "sol é oposto a lua". Em seguida era apresentado um slide onde se encontrava escrito: "Se sol é igual ao dia e lua é igual à noite, e sol é oposto à lua. Dia é oposto à noite? Sol é oposto à noite?". O participante precisava responder que "sim" às duas perguntas e caso o participante errasse o pesquisador perguntava o significado da palavra "oposto" e dizia "oposto significa 'o contrário'". Depois eram repetidas as premissas e as perguntas. Esse processo se repetia até o participante acertar. Posteriormente eram apresentadas as perguntas "'mais' é o oposto de 'menos'" e "'alegre' é o oposto de 'triste'?". Caso o participante errasse ele retornava à apresentação das tentativas anteriores até que obtivesse 100% de acertos.

Posteriormente, as telas do jogo eram apresentadas ao participante. A tela do tutorial continha duas premissas com pseudopalavras e uma pergunta sobre as premissas. As premissas eram: "ZUM é igual a PEG", "PEG é oposto a DUP". Seguidas de uma pergunta: "ZUM é igual a DUP?". O participante então precisava responder que "não". Não havia limite de tempo para que o participante respondesse. Em caso de erro, o pesquisador avisava do erro e repetia que "oposto" significa "ao contrário", repetindo a pergunta. Caso acertasse, perguntava se ainda restavam dúvidas sobre o procedimento. Após a explicação, se iniciava o jogo.

Grupos experimentais e variáveis

Os participantes foram aleatoriamente divididos em quatro grupos de 10 participantes. Os 40 participantes realizaram exatamente as mesmas tentativas, porém com uma diferença na apresentação dos estímulos, o MET. Os estágios "A" variavam os estímulos a cada tentativa, nunca repetindo o estímulo entre cada uma das 12 tentativas do estágio. Já os estágios "B" mantinham o mesmo conjunto de estímulos durante todas as 12 tentativas. Dividimos os grupos em grupo AAAA, ABBA, ABAA e AABA. Assim sendo, descrevendo o que ocorria nos grupos: AAAA variava os estímulos entre todas as tentativas durante todos os quatro estágios; ABBA mantinha o mesmo conjunto de estímulos no estágio 2, e o mesmo conjunto de estímulos em todas as tentativas do estágio 3, variando, porém, entre os estágios 2 e 3; ABAA mantinha o mesmo conjunto de estímulos no estágio 2 e variava nos demais e, por último, o grupo AABA mantinha o conjunto somente nas tentativas do estágio 3. A divisão dos grupos nos quatro estágios de treino pode ser

visualizada na Tabela 3. O estágio 1 foi utilizado como medida da linha de base (sem reforço), enquanto o estágio 4 (final) foi utilizado como medida de comparação do desempenho dos quatro grupos no treino. Esses dois estágios, portanto, se mantiveram com os estímulos variando por tentativa, como no treino tradicional.

Tabela 3

Divisão dos grupos experimentais de acordo com a apresentação dos estímulos (MET)

Grupo experimental	Alteração dos estímulos a cada tentativa			
	1	2	3	4
1. AAAA	Sim	Sim	Sim	Sim
2. ABBA	Sim	Não	Não	Sim
3. ABAA	Sim	Não	Sim	Sim
4. AABA	Sim	Sim	Não	Sim
Estágio do jogo adaptado	1	2	3	4

Nota. O estágio 1 foi utilizado como linha de base idêntica entre todos os quatro grupos e portanto, não foi utilizado o reforço nesse estágio.

A cada tentativa, em caso de erro, a tela era coberta por um quadrado branco e a mensagem "Errado" em negrito, com o botão com a resposta certa se ampliando. Em caso de acerto, a tela era coberta por um quadrado branco e a mensagem "Correto!" em negrito acompanhada de uma música curta, com tom alegre. Ao final de cada estágio de 12 tentativas, independente do desempenho do participante, o pesquisador emitia reforços verbais como "você foi muito bem. Está tudo bem errar às vezes, é normal. Boa sorte no próximo estágio!". Ao final de cada estágio também aparecia uma tela indicando o escore do participante ("você fez X pontos") e dizendo "seu recorde foi de Y pontos" (sendo Y o maior escore que o participante fez em um dos estágios). Ao longo dos quatro estágios a variável dependente (VD) mensurada foi os escores. As variáveis independentes (VI) foram o grupo de treino no qual o participante se encontrava (ABBA, ABAA, AABA e AAAA) e o estágio de treino (1, 2, 3 ou 4).

ANÁLISE DOS DADOS E RESULTADOS

Descritivas

A idade média das 40 crianças foi de 122,38 meses (Desvio-Padrão [DP] = 10,07 mínimo = 111 máximo = 150) \approx 10,19 anos; DP = 0,84. O nível socioeconômico da amostra foi avaliado usando o escore total no Critério Brasil. A média no escore foi de 42,68; DP = 8,74. Para as descritivas gerais, calculamos média (M) e DP divididos por grupo. Os resultados encontram-se na Tabela 4.

Os resultados indicam que os escores não foram superiores em estágios no qual o mesmo conjunto de estímulos foi mantido em relação a estágios em que os estímulos variaram. Esse dado sugere que a inserção dos estágios de variação B não tiveram impactos nos escores dos participantes. Os resultados indicam que em todos os grupos o escore foi superior no estágio 3 em relação ao estágio 1 e 4. Todos os grupos tiveram uma média de escore superior no estágio 4 em relação ao estágio 1.

Tabela 4

Escore médio de cada grupo por estágio

Grupo/estágio	1	2	3	4	
ABBA	M = 5,4	M = 6,1	M = 7,6	M = 6,3	
	DP = 2,32	DP = 2,33	DP = 2,99	DP = 2,63	
ABAA	M = 7,0	M = 6,2	M = 9,0	M = 7,3	
	DP = 2,16	DP = 2,49	DP = 2,54	DP = 2,41	
AABA	M = 5,6	M = 6,2	M = 8,8	M = 7,3	
	DP = 1,27	DP = 1,93	DP = 1,81	DP = 2,21	
AAAA	M = 6,0	M = 6,6	M = 7,7	M = 6,9	
	DP = 2,31	DP = 3,13	DP = 3,20	DP = 2,28	

Análise de covariância comparando quatro grupos experimentais

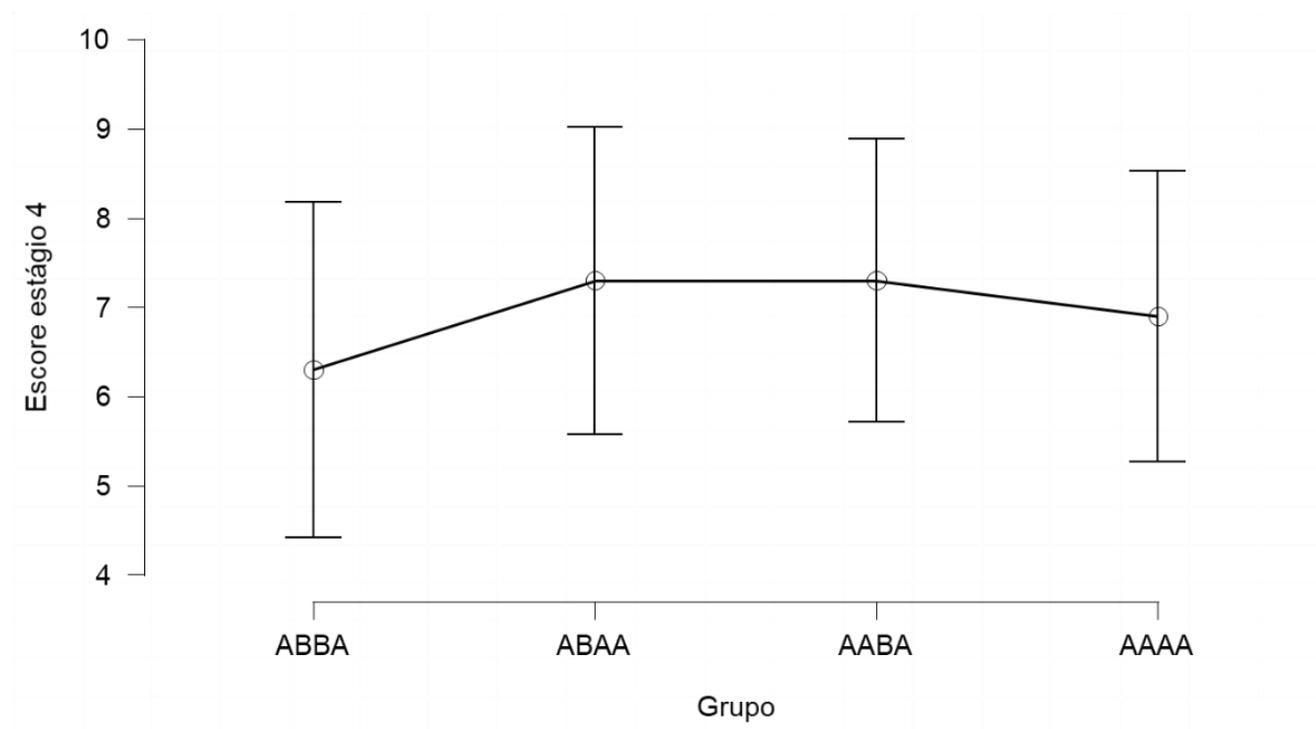
Para verificação dos efeitos do uso de múltiplos estímulos por estágio foi realizada uma análise de covariância (ANCOVA) comparando o escore dos quatro grupos experimentais eliminando a possibilidade de erro por variações dos grupos a partir do estágio 1. Ou seja,

considera-se o desempenho do grupo no estágio 4, mas considera-se também os diferentes pontos de partida (escore no estágio 1) entre os diferentes grupos (ver gráfico da Figura 2).

Não houve diferenças estatisticamente significativas dos escores entre os grupos ($p > 0,05$ em todas as comparações). Os resultados demonstraram que o uso de menos ou mais conjuntos de estímulos em alguns estágios não interferiu no resultado no último estágio nesse estudo.

Figura 2

ANCOVA comparando escore dos quatro grupos no estágio 4.



Correlações de Pearson a partir do Critério Brasil e idade

A fim de verificar se as diferenças no desempenho estavam relacionadas a idade ou a fatores socioeconômicos, uma análise a partir da correlação de Pearson foi realizada comparando os escores nos quatro momentos, o escore no Critério Brasil e a idade dos participantes. Nenhuma das correlações foram significativas ($p < 0,05$) demonstrando que efeitos do treino não estiveram relacionados a idade ou a estimativa de renda familiar. Os resultados das correlações de Pearson encontram-se na Tabela 5.

Tabela 5*Resultados das Correlações de Pearson*

Variável	Escore 1	Escore 2	Escore 3	Escore 4	Total CB	Idade
Total CB	0.031	0.058	-0.053	0.063	—	
Idade	-0.006	0.142	0.044	0.184	-0.18	—

Nota. Os numerais se referem ao momento do treino (estágio 1, 2, 3 ou 4). Total CB = total do escore no Critério Brasil. Não encontramos correlações significativas entre o escore no Critério Brasil, escore nos estágios e idade.

Análise dos parâmetros de treino por tentativa do estágio 1

Considerando que todos os participantes responderam aos mesmos itens e nas mesmas condições na linha de base (estágio 1), utilizamos uma estratégia para investigar quais parâmetros estavam mais associadas à chance de acerto. Primeiro, dividimos as doze possibilidades de tentativas em três tipos principais. Devido à disposição das premissas no estágio 1, em tentativas com distância nodal 1, o participante precisava responder de acordo com uma premissa de igualdade e uma de oposição, ou duas premissas de oposição. Ou seja, ainda que a distância nodal seja a mesma, há tipos diferentes de perguntas derivadas do mesmo parâmetro. Essa diferença é explicitada na Tabela 6.

Para analisar a diferenças entre as probabilidades de acerto de cada um desses tipos de perguntas, dividimos as questões do estágio 1 entre tentativas de igualdade e oposição (IO), tentativas de oposição e oposição (OO) e tentativas de igualdade, oposição e oposição (IOO). Esses três tipos, referem-se, portanto, às relações presentes nas premissas envolvidas pela tentativa. Na Tabela 6 são especificadas as doze possibilidades de tentativas de acordo com esses três tipos (IO, OO ou IOO). Esses tipos de tentativas foram deduzidos a partir dos parâmetros existentes de distância nodal, a fim de pormenorizar a diferença dois tipos de tentativas com a mesma distância nodal. Como demonstrado na Tabela 6, ainda há outros parâmetros que, combinados, geram outras possibilidades de perguntas. Na Tabela 6, entre as perguntas numeradas 1 a 4, ainda que sejam do mesmo tipo, IO, elas ainda variam na direção da questão (A é igual a C versus C é igual a A) e na relação da questão (A é igual a C versus A é oposto a C).

Tabela 6*Tipos de tentativas deduzidas a partir dos parâmetros*

	Tentativas IO	Tentativas OO	Tentativas IOO
Premissas do estágio 1	A é igual a B. B é oposto a C. C é oposto a D.	A é igual a B. B é oposto a C. C é oposto a D.	A é igual a B. B é oposto a C. C é oposto a D.
Possibilidades de tentativas	1. A é igual a C? 2. C é igual a A? 3. A é oposto a C? 4. C é oposto a A?	5. B é igual a D? 6. D é igual a B? 7. B é oposto a D? 8. D é oposto a B?	9. A é igual a D? 10. D é igual a A? 11. A é oposto a D? 12. D é oposto a A?

Nota. O negrito nas premissas é para fins ilustrativos e destaca as premissas em que o participante precisa focar para responder às tentativas. A numeração também é para fins ilustrativos. Exemplo: há quatro possibilidades de tentativas, 1 a 4, de tipo IO. Nessas tentativas, o participante precisa focar nas duas primeiras premissas da tela, de igualdade e oposição, em negrito, para conseguir responder corretamente. Legenda: IO = Igualdade e oposição; OO = Oposição e oposição; IOO = Igualdade, oposição e oposição.

Para a análise realizamos um modelo misto (também chamados de modelos hierárquicos) de regressão logística considerando as probabilidades de acerto nos três diferentes tipos de tentativas. Foram inseridas na análise, ainda, as interações duplas entre direção da questão e relação da questão; relação da questão e tipo de tentativa; direção da questão e tipo de tentativa. E a interação tripla entre direção da questão, tipo de tentativa e relação da questão. Os resultados sugerem que houve efeito significativo do tipo de tentativa; $\chi^2(2) = 12,21$ $p = 0,002$ e uma interação significativa entre o tipo de tentativa e a relação da questão $\chi^2(2) = 22,01$ $p = < 0,001$. Não houve efeitos das demais variáveis ou da interação entre elas.

Para interpretação dos dados, realizamos o post hoc de Bonferroni comparando a interação entre tipos de tentativas e relação da questão. Os resultados são indicados nas Tabelas 7 e 8. Os dados foram apresentados em escala de razão de chance (*odds ratio*, OR). O OR indica a associação entre os parâmetros da tentativa e o acerto em termos de chance. O OR indica quantas vezes a referência tem mais probabilidade de acerto do que a comparação. Nas Tabelas 7 e 8, a categoria de referência está sempre à esquerda. Caso o OR seja menor que 1, só é possível interpretá-lo como menos chances de acerto, não sendo permitida a multiplicação dita anteriormente. Por exemplo: na Tabela 8, primeira linha, é possível verificar um efeito

significativo da interação entre relação da questão e tipo da tentativa nas probabilidades de acerto ($p < 0,001$), o OR é de 6,366. Isso quer dizer que o participante tem 6,366 vezes mais chances de acerto quando a relação da questão é igual e o tipo de questão é IO do que quando a relação da questão é igual e o tipo da questão é IOO. Como o OR foi maior que 1 é possível interpretar o dado a partir dessa multiplicação de chances. Quando o OR é menor que 1, essa multiplicação não é possível na interpretação. No caso da comparação entre tentativas IOO e tentativas OO, o OR foi de 0,5. Esse dado é interpretado como havendo menos chances de acerto em questões do tipo IOO do que em questões do tipo OO (ver Tabela 7, terceira linha). A Tabela 8 demonstra as análises considerando a interação entre tipos de tentativa e relação da questão. Os efeitos da interação são visualizados na Figura 3.

Tabela 7

Post hoc de Bonferroni comparando tipos de tentativas

Comparação	OR	SE	z	P _{bonferroni}
IO > IOO	2.153	0.52	3.174	0.005
IO - OO	1.078	0.254	0.318	1
IOO < OO	0.501	0.12	-2.893	0.011

Nota. OR = Odds ratio, razão de chance. SE = Standard error, erro padrão.

Tabela 8

Post hoc de Bonferroni comparando a interação entre tipos de tentativa e relação da questão

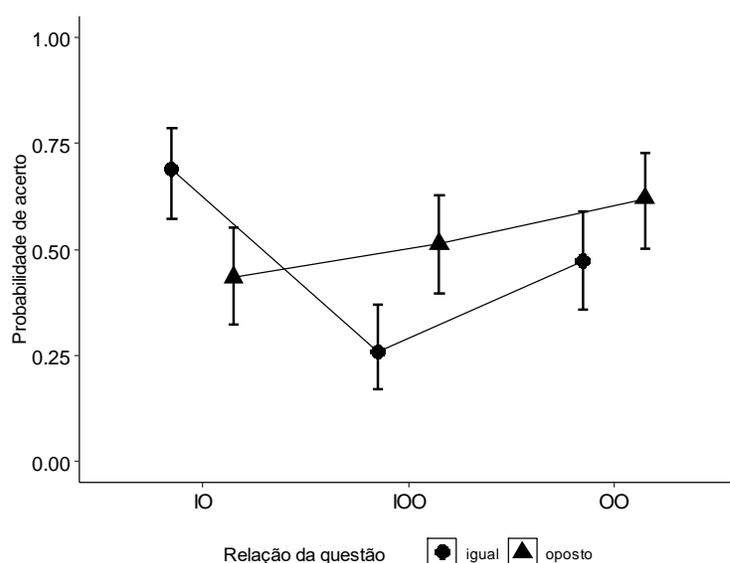
Relação da questão	Tipo de tentativa	Relação da questão	Tipo de tentativa	OR	SE	z	P _{bonferroni}
igual	IO	>	igual IOO	6.366	2.2774	5.174	<.001
igual	IO	-	igual OO	2.469	0.8323	2.68	0.11
igual	IO	>	oposto IO	2.894	0.9794	3.14	0.025
igual	IO	-	oposto IOO	2.107	0.7099	2.212	0.405
igual	IO	-	oposto OO	1.362	0.465	0.905	1
igual	IOO	-	igual OO	0.388	0.1336	-2.75	0.089
igual	IOO	<	oposto IOO	0.331	0.1141	-3.206	0.02
igual	IOO	<	oposto OO	0.214	0.0751	-4.391	<.001
igual	OO	-	oposto OO	0.552	0.1825	-1.797	1
oposto	IO	-	igual IOO	2.2	0.7588	2.286	0.334

oposto	IO	-	igual	OO	0.853	0.2781	-0.488	1
oposto	IO	-	oposto	IOO	0.728	0.2375	-0.973	1
oposto	IO	-	oposto	OO	0.471	0.1563	-2.27	0.348
oposto	IOO	-	igual	OO	1.172	0.381	0.487	1
oposto	IOO	-	oposto	OO	0.646	0.2138	-1.319	1

Nota. OR = *Odds ratio*, razão de chance. SE = *Standard error*, erro padrão. As comparações significativas estão em negrito. A seta indica que a relação da esquerda (referência) tem maior (>) ou menor (<) probabilidade de acerto que a relação comparada à direita.

Figura 3

Probabilidade de acerto a partir da interação entre relação da questão e tipo de tentativa



Os resultados na Figura 3 indicam que a tentativa com maior probabilidade de acerto foi a tentativa do tipo IO, com questões de igualdade, com probabilidade de acerto que chegam até aproximadamente 75%. Demais resultados indicam probabilidades muito baixas de acerto, próximas ou abaixo de 50% o que sugere que os participantes acertaram de forma aleatória. Todas as demais tentativas requisitaram do participante responder de acordo com o uso de duas ou mais relações de oposição simultâneas, respondendo a partir de uma tentativa de tipo IOO ou OO.

DISCUSSÃO

O presente trabalho foi baseado em estudos que indicavam que o treino SMART não era efetivo para todos os participantes (Amd & Roche, 2018; J.Hayes & Stewart, 2016). Nossa investigação foi conduzida a partir de duas hipóteses. A primeira foi que o uso de um mesmo conjunto de estímulos por estágio do treino ocasionaria em um aumento no escore dos participantes; a segunda, que a ordem dos estágios e das tentativas do SMART não seguem uma ordem gradual e progressiva de dificuldade. Assim, o presente estudo teve como propósito verificar se o uso de estímulos repetidos por estágio resultaria em maior escore dos participantes ao final do treino e verificar se a ordem de estágios e de tentativas do SMART aumentam gradualmente o seu nível de dificuldade.

O experimento foi realizado a partir de alterações no MET, na avaliação do escore dos participantes em quatro estágios e das probabilidades de acerto por tentativa de acordo com as variações de parâmetros no mesmo estágio. Os resultados demonstraram evidências que o uso de múltiplos conjuntos de estímulos não interferiu no escore dos participantes e que o SMART parece não demonstrar uma ordem de apresentação nos estágios que considera aumentos graduais de dificuldade. Além disso, encontramos diferenças na probabilidade de acerto de acordo com os parâmetros utilizados para as tentativas, ainda que tentativas do mesmo estágio.

Em relação à hipótese sobre o conjunto de estímulos utilizado por estágio (MET), a diferença de desempenho dos participantes não foi significativa entre grupos que utilizaram o mesmo conjunto de estímulos em alguns estágios em relação ao grupo que utilizou diferentes conjuntos por tentativa. Caso essa hipótese fosse confirmada, os grupos deveriam demonstrar aumentos de desempenho em estágios nos quais o mesmo conjunto de estímulos fosse utilizado. O uso de quatro grupos experimentais visava a demonstração desse efeito somente a partir dos estágios nos quais a manipulação do mesmo conjunto de estímulos fosse utilizada, ou que ao final dos quatro estágios o efeito fosse superior no grupo com mais exposições ao mesmo conjunto de estímulos. Embora esse não tenha sido o caso, consideramos que não se pode descartar completamente a hipótese.

No nosso estudo as tentativas por estágio podem não ter sido suficientes para demonstrar efeitos do mesmo conjunto de estímulos no desempenho dos participantes. Além disso, as

características e tamanho da amostra de participantes também podem não ter sido adequados. A amostra foi de participantes verbalmente competentes e sem nenhum atraso no desenvolvimento ou diagnóstico. Buscou-se adequar o nível de dificuldade do treino a faixa etária da amostra, utilizando estágios mais avançados do treino. Porém, pode ser que, naturalmente, os participantes da amostra já sejam mais orientados às relações dentre estímulos do que aos estímulos em si (Barnes-Holmes et al., 2020; Colbert et al., 2018). Ou seja, o repertório anterior dos participantes já garantiria que eles estivessem mais sob controle das relações entre os estímulos do que sob controle dos estímulos utilizados. Essa é uma hipótese a ser verificada em futuros estudos com populações com menor repertório verbal, como crianças mais jovens ou de desenvolvimento atípico.

Outra investigação que conduzimos paralelamente foi buscar correlacionar a classificação socioeconômica e o desempenho dos participantes nas tentativas. Essa correlação não foi encontrada (ver Tabela 9), dado que corrobora com a interpretação fornecida por Amd & Roche (2018) que indicaram que o SMART poderia ser efetivo independentemente de aspectos socioeconômicos dos participantes.

A segunda hipótese que conduziu os objetivos de nosso estudo foi confirmada: todos os participantes demonstraram desempenhos superiores no estágio do jogo baseado no estágio 28 do SMART em relação aos estágios baseados no 27 e 29. Esses resultados são coerentes com as evidências de que diferentes molduras possuem diferentes níveis de complexidade e que molduras de igualdade são mais básicas que molduras de oposição (O’Hora et al., 2002; Steele & Hayes, 1991). Os dados indicam que um estágio que utiliza duas premissas de oposição e uma de igualdade é mais complexo que um estágio que utiliza duas premissas de igualdade e uma de oposição. Essa evidência corrobora com nossa hipótese de que a ordem dos estágios do SMART não leva em consideração um nível progressivo de dificuldade, e, portanto, pode prejudicar o desempenho dos participantes. Uma sequência potencialmente mais adequada para o aumento progressivo de dificuldade, treinaria os participantes no estágio 28 primeiramente, antes dos estágios 27 e 29. Nosso estudo não verificou o efeito de diferentes ordens de treino no desempenho final dos participantes sendo essa uma sugestão para trabalhos futuros.

Visando sugerir uma ordenação mais adequada dos treinos verificamos, a partir de uma análise de regressão logística com modelo hierárquico, o efeito de diferentes parâmetros nas probabilidades de acerto dos participantes nas tentativas do estágio 1. Os resultados indicaram um

efeito do tipo de tentativa e efeitos da interação entre tipo de tentativa e as relações presentes nas questões. Tentativas do tipo IOO, ou seja, com distância nodal 2, demonstraram ter duas vezes menos chances de acerto que tentativas do tipo IO, com distância nodal 1 (ver Tabela 4). Tentativas do tipo IOO também demonstraram ter menos chances de acerto que tentativas do tipo OO. Os efeitos da distância nodal são coerentes com os reportados na literatura (Bentall et al., 1999; Fields et al., 1990) que indicam que a acurácia no responder reduz em função da distância nodal. Não foram observadas diferenças significativas no desempenho a partir da linearidade das questões. Esse dado é consistente com um experimento anterior que verificou que efeitos da linearidade não interferiram no número de participantes adultos que completaram treinos a partir de molduras relacionais de comparação (Munnelly et al., 2013).

Os nossos dados apontaram que uma das tentativas, ainda que baseada no mesmo estágio do SMART, chega a ter seis vezes mais probabilidade de acerto do que outra do mesmo estágio (ver Tabela 8). Isso implica não só que a ordenação entre os estágios possa estar inadequada, como também a ordenação e apresentação das tentativas intra estágios. O uso de uma dificuldade inadequada para o repertório pode impactar na motivação dos participantes e na queda de desempenho. O equilíbrio entre grau de desafio e desempenho pode ser um fator importante na motivação (Engeser & Rheinberg, 2008).

O uso de uma análise hierárquica trouxe dados para verificação da adequação dos parâmetros pensando em probabilidades de acerto. Este é o primeiro estudo de que temos conhecimento que analisa os parâmetros por tentativa do SMART isoladamente. Baseado exclusivamente em nossos dados, uma ordenação adequada dos exercícios do SMART poderia começar primeiramente por tentativas de menor distância nodal e com questões de igualdade, seguida de tentativas com menor distância nodal e com questões de oposição. Essa disposição pode ser melhor do que uma disposição de forma aleatória dentro do mesmo estágio, mas variações na ordem não foram avaliadas em nosso estudo e precisam continuar a ser testadas em pesquisas futuras.

Uma alternativa seria a utilização de estágios adequados individualmente, de acordo com o desempenho dos participantes. Há estudos que utilizam esse formato automatizado baseados na “Teoria de Resposta ao Item” no que é chamado de testagem adaptativa computadorizada. Nesse tipo de testagem os avanços são ajustados de acordo com o desempenho, participantes que acertam questões com menor probabilidade de acerto, ou seja, mais difíceis, não são testados em questões

mais fáceis (Ricarte, 2013). O uso de testagens computadorizadas tende a reduzir bastante a duração do treino e manter a motivação dos participantes (Lu et al., 2016).

Por último, esse estudo demonstrou as vantagens do uso de um experimento automatizado, mediado por software. Devido às restrições da pandemia de COVID-19 o experimento precisou ser realizado 100% a distância. O uso de um software automatizado também permitiu um maior rigor experimental mantendo constantes a mediação de reforço e o contrabalanceamento aleatório de posições dentre as opções de resposta. Os benefícios do uso de tecnologias automatizadas já foi apresentado por diversos autores e permanece como uma importante frente de avanço nos estudos comportamentais (J. Hayes & Stewart, 2016; Kilroe et al., 2014; Walsh et al., 2014).

CONCLUSÃO

Nosso trabalho focou em algumas variáveis metodológicas e parâmetros que poderiam influenciar no desfecho dos resultados de participantes no treino SMART: o uso de múltiplos conjuntos de estímulos por estágio, as diferenças dos escores em alguns estágios do SMART e a análise da probabilidade de acerto em diferentes tipos de tentativas, além da possível correlação de variáveis socioeconômicas no desempenho. Os resultados corroboram com a hipótese que a ordenação do treino SMART não é eficiente em termos de aumento gradual de dificuldade. Esse aspecto pode impactar no desempenho de alguns participantes no treino e pode ser crítico se pensarmos no uso do SMART para outras populações, como crianças com desenvolvimento atípico. Com base na nossa discussão, futuros trabalhos podem manipular a ordem das tentativas e verificar se os participantes terão melhores escores.

CONCLUSÃO GERAL

Em nosso trabalho tivemos como objetivo geral explorar alternativas procedimentais para o treino SMART. Essa exploração ocorreu no Estudo 1 através de um artigo de revisão sistemática com treinos automatizados baseados na RFT que objetivava encontrar e comparar os artigos empíricos com diferentes treinos; e no Estudo 2 com uma investigação de variáveis que poderiam interferir no escore dos participantes no SMART, incluindo a repetição ou variação de estímulos por tentativa e a comparação do escore e probabilidades de acerto dos participantes a cada tentativa e estágio.

No Estudo 1 apontamos, através dos resultados, que há uma heterogeneidade nos trabalhos com diferentes instrumentos baseados na RFT. Essa heterogeneidade dificultou a comparação dos estudos, mas trouxe recomendações para uma agenda de pesquisa de treinos relacionais a partir da exploração de diversas variáveis como a aplicações dos treinos com molduras relacionais diferentes, amostras mais diversas, o uso de diferentes protocolos e estruturas de treino e a variação do tipo dos estímulos utilizados tanto no treino quanto nas opções de resposta. Futuros trabalhos podem investigar essas diferentes variáveis e buscar comparações entre diferentes instrumentos, desde que controlando alguns fatores, como por exemplo a amostra, ao realizarem a comparação.

Além disso, alguns dos trabalhos encontrados a partir do Estudo 1 nos trouxeram questionamentos a respeito da estrutura de treino do SMART. Dois trabalhos com o instrumento RCP (May et al., 2017; Walsh et al., 2014) utilizaram um procedimento diferente de MET do SMART, não variando estímulos tentativa a tentativa como no SMART e, ao invés disso, reforçando os participantes ordenadamente em relações de linha de base e em respostas de implicação combinatória em um mesmo conjunto de estímulos antes de realizarem o procedimento com um novo conjunto de exemplares. Também encontramos em um dos trabalhos com o TARPA a utilização de uma análise hierárquica para verificar a melhor ordenação dos treinos nesse instrumento (Moran et al., 2014). Os três trabalhos citados (May et al., 2017; Moran et al., 2014; Walsh et al., 2014) nos permitiram levantar perguntas em busca de alternativas que possivelmente aumentassem o número de participantes que completam o treino SMART, posto que dados anteriores indicaram que nem todos os participantes avançavam sequer do primeiro estágio (Amd & Roche, 2018). Não encontramos outros trabalhos com o SMART que demonstrassem como em Amd & Roche (2018), quais participantes não conseguiram avançar de certos estágios o que dificultou avaliar quais variáveis impactam no desempenho dos participantes. Nos questionamos

se o uso de estímulos repetidos entre tentativas aumentaria o escore dos participantes tanto no estágio mensurado quanto ao final do treino e se o treino SMART era organizado a partir de uma ordem gradual de dificuldade. Nossos dados não apontaram diferenças nos escores comparando grupos com o uso de estímulos repetidos com grupos que variaram os estímulos tentativa a tentativa. Todavia, demonstramos evidências que o SMART não é organizado de forma gradual de dificuldade nos estágios e que mesmo tentativas dentro de um mesmo estágio possuem menor possibilidade de acerto que outras. A partir dos nossos achados, pesquisas futuras podem buscar seguir a ordem sugerida de treino na discussão, para os estágios e tentativas por estágio, e comparar o escore dos participantes na ordem sugerida em contraposição a uma ordem aleatória de tentativas, como ocorre no SMART tradicional.

Nosso trabalho foi o primeiro trabalho encontrado que buscou comparar treinos automatizados baseados na RFT e o primeiro trabalho encontrado a tentar propor alterações metodológicas para o SMART e realizar uma investigação empírica do grau de dificuldade das tentativas nesse protocolo. Esperamos que, a partir dos resultados dos dois estudos, de revisão e empírico, pesquisadores interessados em treinos relacionais possam investigar uma série de perguntas ainda não exploradas, mas apontadas a partir dos resultados dos dois trabalhos. Uma ordem gradual de dificuldade do SMART aumentaria o escore dos participantes? O uso de estímulos repetidos impactaria nos escores em uma amostra de desenvolvimento atípico ou mais jovem? O uso de MET como no SMART, nos demais instrumentos de treino, também demonstraria aumentos nos escores dos participantes em testes psicométricos e de habilidades acadêmicas? Qual protocolo de treino (SMART, TARPA, RCP, REP ou T-IRAP) impactaria mais nos escores de testes psicométricos e habilidades acadêmicas, caso fosse utilizado o MET? Essa e mais uma série de outras perguntas podem ser levantadas a partir dos nossos resultados e podem ser investigadas por futuros pesquisadores.

REFERÊNCIAS

- Associação Brasileira de Empresas de Pesquisa (2021). *Critério Brasil*.
- Amd, M., & Roche, B. (2018). Assessing the Effects of a Relational Training Intervention on Fluid Intelligence Among a Sample of Socially Disadvantaged Children in Bangladesh. *The Psychological Record, 68*(2), 141–149. <https://doi.org/10.1007/s40732-018-0273-4>
- Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2020). Updating RFT (More Field than Frame) and its Implications for Process-based Therapy. *The Psychological Record, 70*(4), 605–624. <https://doi.org/10.1007/s40732-019-00372-3>
- Bentall, R. P., Jones, R. M., & Dickins, D. W. (1999). Errors and Response Latencies as a Function of Nodal Distance in 5-Member Equivalence Classes. *The Psychological Record, 49*(1), 93–115. <https://doi.org/10.1007/BF03395309>
- Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences, 47*, 222–235. <https://doi.org/10.1016/j.lindif.2016.03.001>
- Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *The Psychological Record, 61*, 173-198.
- Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*(3), 158–172. <https://doi.org/10.1007/s11031-008-9102-4>
- Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). THE EFFECTS OF NODALITY ON THE FORMATION OF EQUIVALENCE CLASSES. *Journal of the Experimental Analysis of Behavior, 53*(3), 345–358. <https://doi.org/10.1901/jeab.1990.53-345>
- Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *British Journal of Educational Psychology, 86*(3), 397–411. <https://doi.org/10.1111/bjep.12114>
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Orgs.). (2001). *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. Springer US. <https://doi.org/10.1007/b108413>
- Holth, P. (2017). Multiple Exemplar Training: Some Strengths and Limitations. *The Behavior Analyst, 40*(1), 225–241. <https://doi.org/10.1007/s40614-017-0083-z>
- Kilroe, H., Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2014). Using the T-IRAP interactive computer program and applied behavior analysis to teach relational responding in children with autism. *Behavioral Development Bulletin, 19*(2), 60–80. <https://doi.org/10.1037/h0100578>
- LaFrance, D. L., & Tarbox, J. (2019). The importance of multiple exemplar instruction in the establishment of novel verbal behavior. *Journal of Applied Behavior Analysis, jaba.611*. <https://doi.org/10.1002/jaba.611>
- Lu, H., Hu, Y. P., & Gao, J. J. (2016). The effects of computer self-efficacy, training satisfaction and test anxiety on attitude and performance in computerized adaptive testing. *Computers & Education, 100*, 45-55.
- McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Relational Operant Skills Training Increases Standardized Matrices Scores in Adolescents: A Stratified Active-Controlled Trial. *Journal of Behavioral Education. https://doi.org/10.1007/s10864-020-09399-x*
- McLoughlin, S., Tyndall, I., Pereira, A., & Mulhern, T. (2021). Non-verbal IQ Gains from Relational Operant Training Explain Variance in Educational Attainment: An Active-

- Controlled Feasibility Study. *Journal of Cognitive Enhancement*, 5(1), 35–50. <https://doi.org/10.1007/s41465-020-00187-z>
- Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational ability and language performance in children with autism spectrum disorders and typically developing children: A further test of the TARPA protocol. *The Psychological Record*, 64(2), 233–251.
- Munnelly, A., Freegard, G., & Dymond, S. (2013). Constructing Relational Sentences: Establishing Arbitrarily Applicable Comparative Relations with the Relational Completion Procedure. *The Psychological Record*, 63(4), 751–768. <https://doi.org/10.11133/j.tpr.2013.63.4.004>
- O’Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record*, 52, 51–76.
- O’Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived Relational Responding and Performance on Verbal Subtests of the WAIS-III. *The Psychological Record*, 55(1), 155–175. <https://doi.org/10.1007/BF03395504>
- O’Hora, D., Peláez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal Relations and Intelligence: Correlating Relational Performance With Performance on the Wais-III. *The Psychological Record*, 58(4), 569–584. <https://doi.org/10.1007/BF03395638>
- Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, 8(4), 434–447. <https://doi.org/10.1037/1082-989X.8.4.434>
- O’Toole, C., & Barnes-Holmes, D. (2009). Three Chronometric Indices of Relational Responding as Predictors of Performance on a Brief Intelligence Test: The Importance of Relational Flexibility. *The Psychological Record*, 59(1), 119–132. <https://doi.org/10.1007/BF03395652>
- Ricarte, T. A. M. (2013). *Teste adaptativo computadorizado nas avaliações educacionais e psicológicas* [Mestrado em Ciências de Computação e Matemática Computacional, Universidade de São Paulo]. <https://doi.org/10.11606/D.55.2013.tde-16052013-143315>
- Steele, D., & Hayes, S. C. (1991). STIMULUS EQUIVALENCE AND ARBITRARILY APPLICABLE RELATIONAL RESPONDING. *Journal of the Experimental Analysis of Behavior*, 56(3), 519–555. <https://doi.org/10.1901/jeab.1991.56-519>
- Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational frame theory, mathematical, and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology & Psychological Therapy*, 16(2), 141–155.
- Walsh, S., Horgan, J., May, R. J., Dymond, S., & Whelan, R. (2014). Facilitating relational framing in children and individuals with developmental delay using the relational completion procedure: FACILITATING RELATIONAL FRAMING. *Journal of the Experimental Analysis of Behavior*, 101(1), 51–60. <https://doi.org/10.1002/jeab.66>

Anexo A – Termo de Consentimento Livre e Esclarecido

O Sr. (a) e seu dependente legal estão sendo convidados (as) como voluntários (as) a participar da pesquisa “É POSSÍVEL ABREVIAR OS TREINOS DE HABILIDADES RELACIONAIS? COMPARAÇÃO DE DOIS TREINOS UTILIZANDO MOLDURAS DE COORDENAÇÃO E OPOSIÇÃO” coordenado pela profa. Dra. Thais Porlan de Oliveira do Departamento de Psicologia da Universidade Federal de Minas Gerais (UFMG). Serão utilizados em pesquisa os resultados nas atividades psicológicas explicada a seguir. A utilização desses dados está vinculada somente a publicações acadêmicas científicas, sendo garantidas todas as medidas de sigilo para a não identificação do (a) Sr. (a) e da criança convidada. Pedimos a sua autorização como responsável legal da criança para a utilização desses dados sigilosos da mesma obtidos em pesquisa científica.

Essa pesquisa tem como objetivo a adaptação de um procedimento de treino com atividades lógicas. Esse procedimento se baseia nos estudos que apontam como esse tipo de treino pode ajudar na compreensão do desenvolvimento da linguagem, raciocínio lógico e abstrato. Os participantes convidados serão crianças de 9 a 11 anos que serão divididos em quatro grupos experimentais e um controle, que receberão treinos semelhantes, com os mesmos exercícios, sendo a principal diferença a ordem de apresentação dos estágios do treino. Durante o estudo será realizada uma chamada de vídeo via aplicativo ZOOM, a ser agendada pelo pesquisador de acordo com a disponibilidade do(a) cuidador(a) e da criança. A criança necessita de um computador com câmera e a disponibilidade de 30 minutos para a realização da atividade, que será realizada em somente um encontro. São destinados 6 minutos à apresentação do estudo e retirada de dúvidas e 24 minutos de aplicação do procedimento. Para todos os grupos de participantes o treino de 24 minutos é dividido em 4 estágios de 6 minutos cada. Em todos os grupos, há 48 exercícios totais, 12 por estágio. A dificuldade de cada tentativa varia, podendo ser mais simples ou difícil, porém sempre adequada à idade dos participantes.

A coleta de dados será realizada pelo estudante de mestrado orientado para o projeto de pesquisa, Daniel Ezequiel Pinto ou uma das duas estudantes de psicologia da UFMG, devidamente treinadas para tal fim, Diuliana Nader e Vanessa Torchetti Araújo Rodrigues. Para registro dos dados e análises posteriores as chamadas de vídeo serão gravadas e armazenadas. Para participar deste estudo, o menor sob sua responsabilidade não terá nenhum custo, nem receberá nenhum

tipo de pagamento, em nenhuma circunstância. Ele será esclarecido em todos os aspectos que desejar e estará livre para aceitar ou recusar-se a participar. Você, como responsável pelo menor, poderá retirar seu consentimento ou interromper a participação dele a qualquer momento. A participação dele é voluntária e a recusa em participar não acarretará qualquer penalidade ou modificação na forma em que é atendido pelo pesquisador.

Esse estudo apresenta pequenos riscos físicos e psicológicos para os participantes. Os riscos psicológicos incluem os mesmos riscos existentes em atividades rotineiras como assistir às aulas e realizar atividades curriculares ou outras, no computador: cansaço relativo às tentativas que demandam um esforço cognitivo e possíveis constrangimentos ou frustrações mediante a realização de exercícios mais difíceis. Os riscos físicos são os mesmos de escrever, sentar-se diante de um computador ou assistir a uma aula, como algum desconforto de ordem física ou ergonômica. Caso a criança sintasse cansada, o pesquisador poderá interromper ou até mesmo suspender a avaliação.

É pedido aos cuidadores que preencham um questionário de avaliação socioeconômica (Critério Brasil) e isso pode causar algum constrangimento, mas essas respostas são de acesso exclusivo dos pesquisadores. Ao final da pesquisa, o (a) Sr. (a) receberá um relatório descrevendo os principais resultados desse trabalho, contendo, de forma clara e objetiva, os resultados e explicações dos instrumentos aqui utilizados. As publicações resultantes desse trabalho preservarão o anonimato dos participantes. Qualquer dado apresentado por meio de artigos científicos ou dissertações não incluirá nenhuma informação que possa identificar qualquer participante.

Uma via deste termo de consentimento e do Critério Brasil preenchidos também será enviada para seu e-mail. Os dados, materiais e instrumentos utilizados na pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos e após esse tempo serão destruídos. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, atendendo a legislação brasileira (Resoluções Nº 466/12; 441/11 e a Portaria 2.201 do Conselho Nacional de Saúde e suas complementares), utilizando as informações somente para fins acadêmicos e científicos.

Caso tenha alguma dúvida de ordem ética, entrar em contato com o Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, Rua Av. Antônio Carlos, 6627, Unidade

Administrativa II - 2º andar (Sala 2005), Campus Pampulha, 31270-901- Belo Horizonte (MG),
Fone:(31) 3409-4592. Email coep@prpq.ufmg.br

Eu, _____ (seu nome), portador do documento de
Identidade _____, responsável pelo menor

_____ (nome do seu dependente legal), fui informado dos
objetivos do presente estudo de maneira clara e detalhada e esclareci minhas dúvidas. Sei que a
qualquer momento poderei solicitar novas informações junto ao pesquisador responsável listado
abaixo ou ao Comitê de Ética em Pesquisa. Tenho ciência que posso retirar a participação do
meu dependente legal a qualquer momento, se assim desejar.

Contato dos pesquisadores:

Profa. Dra. Thais Porlan de Oliveira

Pesquisadora responsável pelo projeto

Telefone: (31) 9386-3672

porlan@ufmg.com

Daniel Ezequiel Pinto -

Pesquisador responsável pelo projeto

Telefone: (31) 98413-0647

dezequielp@gmail.com

Anexo B– Termo de Assentimento Livre e Esclarecido

Você está sendo convidado(a) a fazer parte de uma pesquisa chamada “É POSSÍVEL ABREVIAR OS TREINOS DE HABILIDADES RELACIONAIS? COMPARAÇÃO DE DOIS TREINOS UTILIZANDO MOLDURAS DE COORDENAÇÃO E OPOSIÇÃO”. Esta pesquisa quer entender melhor a forma de pensar e se comunicar das pessoas. Para isso, estamos usando atividades no computador para nos ajudar. Durante essas atividades você responderá a algumas perguntas em um jogo no computador enquanto faz uma videochamada com o pesquisador, com a câmera ligada. Todo o tempo você receberá uma explicação sobre as atividades e a videochamada estará sendo gravada, para uso só na nossa pesquisa. Você só precisa participar do nosso estudo se quiser. Se não quiser participar, está tudo bem, é só nos dizer.

Além disso, esse estudo apresenta alguns riscos e é importante que você entenda eles, antes de decidir participar. Terá perguntas e desafios de diferentes tipos e dificuldades. Pode ser que as atividades te cansem por ter de pensar demais, mas você pode parar a hora que quiser. Basta pedir ao pesquisador para deixar as atividades ou fazer uma pausa, sem nenhum problema. Você pode ficar chateado se não conseguir responder algumas perguntas ou incomodado por ter de ficar sentado e concentrado enquanto faz as atividades. Caso não goste de algo relacionado ao nosso trabalho, você pode conversar com o pesquisador, ou com seu cuidador/a e vamos te ajudar. O tempo máximo do estudo é de 30 minutos e depois disso você não precisará responder mais perguntas. Neste tempo você receberá no início explicações do pesquisador ou da pesquisadora e depois iniciará as tentativas no computador.

Para participar deste estudo o seu responsável precisa autorizar a partir do formulário que enviamos antes, deixando você participar. Eles também podem escolher que você não participe mais do estudo a qualquer momento. O seu nome não aparecerá em nenhum trabalho da nossa equipe de pesquisa. Ou seja, será segredo para qualquer pessoa fora da equipe e dos seus cuidadores que você participou do estudo e como você foi nas tentativas. Tudo que você responder ficará guardado conosco por um período de 5 anos e após esse tempo será destruído. Uma via desse documento assinado ficará com você e outra com o pesquisador.

Caso tenha alguma dúvida sobre essa pesquisa, você ou seus responsáveis podem entrar em contato com o Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, Rua Av. Antônio Carlos, 6627, Unidade Administrativa II - 2º andar (Sala 2005), Campus Pampulha, 31270-901- Belo Horizonte (MG), Fone:(31) 3409-4592.

Também pode entrar em contato com os pesquisadores:

Profa. Dra. Thais Porlan de Oliveira

Pesquisadora responsável pelo projeto

Telefone: (31) 9386-3672

porlan@ufmg.com

Daniel Ezequiel Pinto -

Pesquisador responsável pelo projeto

Telefone: (31) 98413-0647

dezequielp@gmail.com

Eu entendi os objetivos da pesquisa e como ela funciona e sei que a qualquer momento poderei tirar dúvidas com um dos pesquisadores responsáveis com o nome escrito logo abaixo dessa folha, ou ainda com meus pais ou o Comitê de Ética em Pesquisa. Tive a oportunidade de ler e tirar as minhas dúvidas.

Sim Não

Seu nome: