

Efeitos de Videomodelação por Si Mesmo e por Par Sobre Respostas Textuais em Crianças com TEA¹

(Effects of Self and Peer Video Modeling on Textual Responses in Children with ASD)

Maria Vivianne Pereira dos Santos^{*,**} e Angelo Augusto Silva Sampaio^{*,2}

^{*}Universidade Federal do Vale do São Francisco

^{**}Crescer – Centro de Especialidades Terapêuticas

(Brasil)

Resumo

A videomodelação consiste na exibição de vídeos retratando um comportamento para instalar ou modificar o respectivo comportamento do espectador, que, após assistir ao vídeo, é solicitado a imitá-lo. O presente estudo comparou videomodelação por si mesmo e videomodelação por par para o ensino de respostas textuais sob controle de consoantes do alfabeto português em três crianças diagnosticadas com Transtorno do Espectro Autista (TEA). Para isso, foi utilizado um delineamento misto de linha de base múltipla entre participantes e de tratamentos alternados. Alguns cuidados no controle experimental foram acrescentados em relação a trabalhos anteriores: as consequências para a emissão de respostas corretas e incorretas nas condições de linha de base e de tratamento foram idênticas; os pares utilizados foram fisicamente semelhantes aos participantes; e incluiu-se a avaliação de preferência dos participantes pelos vídeos. Os resultados indicaram efeitos semelhantes do ensino em ambas as condições de videomodelação, sendo que dois dos três participantes aprenderam igualmente nas duas condições e um participante atingiu o critério apenas na videomodelação por par. As evidências sugerem efeitos positivos advindos do tipo de intervenção avaliado e a necessidade de futuras pesquisas avaliarem o ensino de outras habilidades.

1 Trabalho baseado na dissertação de mestrado da primeira autora, defendida junto ao Mestrado em Psicologia da Univasf, sob orientação do segundo autor. Os autores agradecem a Adriano Alves Barboza, Bruna Colombo dos Santos e Christian Vichi pelas sugestões a versões prévias do manuscrito; e a Miriam Vitória Fernandes Tavares, pelo auxílio durante a coleta de dados.

2 Endereço para correspondência: Angelo Augusto Silva Sampaio, Colegiado de Psicologia, Universidade Federal do Vale do São Francisco, Av. José de Sá Maniçoba, S/N, Centro, Petrolina, PE, 56304-205, Brasil. E-mail: angelo.sampaio@univasf.edu.br

Palavras-chave: videomodelação, modelação, aprendizagem observacional, imitação, autismo, respostas textuais, comportamento verbal, leitura

Abstract

Video modeling—an empirically supported intervention—consists of exhibiting videos portraying a behavior to install or modify the respective behavior of the viewer, who, after watching the video, is asked to imitate it. The present study compared self video modeling and peer video modeling in teaching textual responses under the control of consonants of the Portuguese alphabet to three children diagnosed with Autism Spectrum Disorder (ASD). A mixed between-participant multiple baseline and alternating treatment design was employed. Two boys and one girl, between four and seven years old, were exposed to: a preference assessment to determine potentially reinforcing items, a textual response baseline phase with differential reinforcement but no prompts or video modeling, a post-video elaboration probe, a video preference assessment, and a video modeling phase (with self and peer video conditions). Experimental control was improved in relation to previous studies: consequences for the emission of correct and incorrect responses in all phases were identical; the peers were physically similar to the participants; and the participant's preference for videos was evaluated. The results indicated similar effects of both video modeling conditions on correct textual responses for two of the three participants. The third participant reached the performance criterion in the self video modeling condition after more sessions than the other participants, and stabilized his correct responses in 60% in the peer video modeling. Overall, both self and peer video modeling appear effective in teaching textual responses. The slightly better results in the self video modeling condition can be attributed to the emission of echoic responses related to the consonants employed in that condition during the elaboration of the videos. All participants preferred the peer video, what may be related to the use of peers who were physically similar to the participants. Evidence suggests positive effects arising from the type of intervention evaluated and the need for future research to evaluate the teaching of other skills.

Keywords: video modeling, modeling, observational learning, imitation, autism, textual responses, verbal behavior, reading

Segundo o Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5-TR, ver *American Psychiatric Association - APA*, 2023), pessoas com Transtorno do Espectro Autista (TEA) apresentam prejuízos nas áreas de comunicação e interação social, além de comportamentos repetitivos e interesses restritos. Nesse sentido, é importante a utilização de intervenções empiricamente sustentadas voltadas a essas demandas, conforme recomendado por órgãos como a Organização Pan-Americana da Saúde e a Organização Mundial de Saúde (2023; ver também Steinbrenner et al., 2020). Steinbrenner et al. (2020) identificaram 28 intervenções empiricamente sustentadas para o TEA, sendo 24 originárias da Análise do Comportamento Aplicada (do termo em inglês *Applied Behavior Analysis*, ABA, ver Baer et al., 1968). Uma delas é a videomodelação, a qual apresenta evidências

significativas para o ensino de habilidades acadêmicas e pré-acadêmicas, de comunicação, adaptativas e de autoajuda, motoras, dentre outras (Bellini & Akullian, 2007). A videomodelação consiste na exibição de vídeos retratando um comportamento, com o objetivo de instalar ou modificar o respectivo comportamento do espectador, que, após assistir ao vídeo, é solicitado a imitar o comportamento que foi apresentado (Bellini & Akullian, 2007).

Estudos têm demonstrado a eficácia da videomodelação para o ensino de: leitura fluente (Decker & Buggey, 2012); aplicação de programas de ensino por cuidadores e profissionais (e.g., Barboza et al., 2015; Varella & de Souza, 2018); e habilidades sociais e de comunicação (e.g., Acar et al., 2016; Charlop et al., 2018). O uso de videomodelação tem como vantagens: (1) o menor custo financeiro, porque sua aplicação pode ser realizada diversas vezes em vários ambientes (e.g., casa, escola, clínica) e em qualquer momento, sem a presença de terapeutas (Brasilense et al., 2018; Marcus & Wilder, 2009); (2) ser possivelmente menos aversivo para a população com TEA do que o contato face a face com outras pessoas (Brasilense et al., 2018); (3) a padronização do procedimento para ensinar a habilidade-alvo (Brasilense et al., 2018); e (4) ser uma intervenção socialmente mais aceitável para algumas famílias, o que pode aumentar o envolvimento direto de pais/cuidadores no treino (Barboza et al., 2015; Marcus & Wilder, 2009; Varella & Souza, 2018). Prater et al. (2012) também sugeriram, ao considerar aspectos da prática profissional, que a videomodelação por pares pode representar maior facilidade na produção dos materiais de ensino do que a modelação ao vivo, visto que um único vídeo poderia ser utilizado com diversos clientes.

A videomodelação pode ser realizada utilizando como modelo o próprio indivíduo a ser ensinado—a chamada videomodelação por si mesmo (*self video modeling*)—ou alguma outra pessoa—a chamada videomodelação por par (*peer video modeling*). Ainda não há na literatura clareza em relação a qual modalidade de videomodelação é a mais eficaz. Sherer et al. (2001) compararam a videomodelação por par e por si mesmo para o ensino de habilidades de conversação em cinco crianças diagnosticadas com TEA. Cada criança assistiu a um vídeo no qual ela era retratada participando de uma conversa ou a um vídeo em que um par era retratado participando de uma conversa. Os pais receberam um cronograma de exibição e foram instruídos a mostrar o vídeo três vezes antes da criança dormir, à noite. A programação alternava dias entre o vídeo de si mesmo e o de par. Os participantes não assistiram aos vídeos em nenhum outro momento, e no dia seguinte, após assistir ao vídeo, o pesquisador fazia as perguntas relativas ao vídeo assistido na noite anterior. Essa fase continuou até que as respostas de conversação do participante atingissem 100% de acertos ou se mantivessem estáveis ao longo das sessões durante várias semanas. Os resultados não mostraram nenhuma diferença significativa no desempenho de três crianças em ambas as condições, enquanto uma apresentou melhor desempenho assistindo ao vídeo de si e outra criança apresentou melhor desempenho ao assistir o vídeo do par.

Decker e Buggey (2012) também compararam videomodelação por par e por si mesmo, mas com crianças com diagnóstico de deficiência de aprendizagem específica. O objetivo do estudo foi comparar os efeitos dos dois tipos de vide-

modelação no ensino da leitura fluente (i.e., leitura com precisão e velocidade) de nove estudantes, seis deles do Ensino Fundamental regular e três escolhidos aleatoriamente de uma sala de educação especial. Os estudantes foram divididos aleatoriamente em três grupos: videomodelação por par, videomodelação por si mesmo e grupo controle. O delineamento, com linha de base múltipla entre participantes, contou com três fases: linha de base, intervenção e manutenção. Na linha de base, foi avaliada a fluência e a compreensão da leitura pela criança, por meio de sondas com trechos de 150 a 500 palavras retirados dos livros que os participantes estavam lendo. Durante a fase de intervenção, os participantes assistiam 10 vezes ao vídeo e após quatro sondas a fase era encerrada. Na manutenção, havia sondas de leitura, sem a presença da intervenção. Os resultados indicaram que a fluência de leitura aumentou com os dois tipos de videomodelação, contudo, no grupo controle houve um aumento lento no desempenho na leitura de dois participantes. Além disso, o ensino da fluência de leitura no grupo de videomodelação por si mesmo foi imediato e considerável.

O estudo de Decker e Buggey (2012) indica que uma das habilidades que pode ser ensinada por meio de videomodelação é o comportamento textual, um dos operantes verbais descritos por Skinner (1957/1992), o qual envolve uma classe de respostas (geralmente) vocais sob controle de estímulos visuais (i.e., texto) mantida por reforçadores generalizados (ver também Barros, 2003; Catania, 1999). Por exemplo, uma criança que na presença da palavra escrita “BOLA”, emite a resposta vocal “bola”, e em seguida a professora elogia afirmando “Muito bem!”.

Marcus e Wilder (2009), diante da escassez de estudos sobre o ensino de operantes textuais por videomodelação, compararam o ensino desse operante verbal com videomodelação por si mesmo e videomodelação por par para três crianças com TEA. As respostas textuais ensinadas aos participantes deveriam ficar sob controle de letras gregas e árabes, a fim de evitar uma história de aprendizagem prévia. No vídeo de si mesmo, inicialmente a pesquisadora apresentava uma letra e perguntava à criança qual letra era aquela. O vídeo era editado de modo que após a pesquisadora perguntar pelo nome da letra, a criança aparecia emitindo a resposta vocal. Entretanto, a pesquisadora filmara a própria criança emitindo a resposta-alvo após a pesquisadora ter fornecido um modelo ao vivo da resposta vocal, sem que a criança tivesse tido contato com a letra escrita. No vídeo de pares, o procedimento era semelhante, mas conduzido com outras crianças, colegas dos participantes. A família, sob supervisão da pesquisadora, foi responsável por mostrar o respectivo vídeo (de si mesmo ou de par) para os participantes três vezes ao dia, por dois dias consecutivos. Depois, na sessão experimental, a pesquisadora exibiu o vídeo novamente e logo solicitou a resposta do participante, apresentando e perguntando que letra era aquela, em cinco tentativas por sessão, uma para cada letra.

Marcus e Wilder (2009) concluíram que a videomodelação por si mesmo se mostrou mais efetiva para o ensino de respostas textuais do que a videomodelação por par, visto que com o vídeo de si mesmo os três participantes atingiram o critério de 80% de acerto em três sessões consecutivas, enquanto no vídeo de par somente um dos três participantes atingiu o critério, sugerindo que as diferenças físicas entre os pares e os participantes possam ter influenciado nesse resultado. Entretanto, os

autores apontaram algumas limitações no seu estudo, recomendando modificações metodológicas para pesquisas futuras. Em primeiro lugar, durante o procedimento os pais relataram que as crianças solicitaram assistir mais aos vídeos de si mesmo do que aos vídeos de pares. Isso sugere que o melhor desempenho com a videomodelação de si mesmo poderia se dever à preferência por esse vídeo. Para começar a avaliar essa hipótese, Marcus e Wilder recomendaram investigar objetivamente a preferência dos participantes em relação aos tipos de vídeos. Outra sugestão seria utilizar modelos o mais semelhantes o possível dos participantes para a gravação do vídeo de pares, pois autores como Bandura et al. (1961) observaram que é maior a probabilidade de as pessoas imitarem o comportamento de modelos semelhantes a elas.

Marcus e Wilder (2009) sugeriram ainda que pesquisas futuras sobre videomodelação programassem condições de linha de base e de intervenção idênticas, pois na linha de base do seu estudo não houve *feedback* para o desempenho do participante, ao contrário do que ocorreu durante a intervenção, o que eles interpretaram como uma variável estranha. Ainda mencionaram que seria interessante replicar o estudo deles utilizando estímulos educacionalmente relevantes (e.g., no Brasil, letras do alfabeto da língua portuguesa).

Considerando as vantagens do uso de videomodelação no ensino (Barboza et al., 2015; Marcus & Wilder, 2009; Varella & Souza, 2018); a escassez de estudos sobre o operante textual (Guerra & Almeida-Verdu, 2016; Martone & Santos-Carvalho, 2012); a necessidade de refinamentos metodológicos em estudos comparando videomodelação por si mesmo e videomodelação por par (Marcus & Wilder, 2009); e a relevância social (Baer et al., 1968) aumentada do uso de estímulos educacionalmente relevantes, a presente pesquisa pretendeu ampliar o estudo de procedimentos baseados em videomodelação por si mesmo e videomodelação por par para o ensino de respostas textuais. Para isso, propôs modificar os aspectos metodológicos sugeridos por Marcus e Wilder (2009): o controle das consequências para a emissão de respostas textuais corretas e incorretas nas condições de linha de base e de tratamento, a fim de as tornarem idênticas; o emprego de modelos o mais semelhantes o possível aos participantes; a introdução de uma avaliação da preferência dos participantes pelos vídeos; e a utilização de estímulos educacionalmente relevantes, no caso, consoantes do alfabeto português brasileiro. O objetivo do estudo, portanto, foi comparar os efeitos da videomodelação por si mesmo e por pares fisicamente semelhantes sobre a emissão de respostas textuais de consoantes do alfabeto português, aperfeiçoando o controle experimental do estudo de Marcus e Wilder.

Método

Participantes

Foram recrutadas para a pesquisa seis crianças idades entre quatro e sete anos, com TEA, em atendimento especializado em ABA para TEA em uma clínica de Petrolina-PE, com de 10 hr a 30 hr de aplicação semanal, e que atendiam aos

seguintes pré-requisitos comportamentais: rastreio visual, emissão de respostas ecoicas no mínimo de sílabas, imitação generalizada e seguimento de instruções. Para avaliar os pré-requisitos foi consultado o prontuário dos participantes na própria clínica, no qual constava a avaliação feita por uma mestra em ABA, supervisora dos casos, utilizando o Programa de Avaliação e Nivelamento de Marcos de Comportamento Verbal (*Verbal Behavior Milestones Assessment and Placement Program*; VB-MAPP; Sundberg, 2014).

O recrutamento dos participantes foi feito mediante a divulgação da pesquisa através de panfletos na referida clínica, onde constava telefone e e-mail para contato da pesquisadora e as demais informações necessárias para se voluntariar à pesquisa. Foi informado para os responsáveis dos possíveis participantes como aconteceria o procedimento e que, durante a pesquisa, o responsável não deveria expor o participante a situações na qual esse pudesse emitir respostas textuais sobre consoantes e se houvesse alguma situação relacionada, que deveria informar à pesquisadora. Os responsáveis que aceitaram participar do estudo, assinaram um Termo de Consentimento Livre e Esclarecido (TCLE). Além disso, foi também necessário o assentimento dos próprios participantes. Como eles não possuíam o repertório verbal necessário para consentir explicitamente com a participação, eles foram previamente instruídos a partir de uma linguagem acessível e o assentimento foi verificado a partir da observação do comportamento delas ao iniciar o procedimento (não chorar ou apresentar recusas). A presente pesquisa foi aprovada por um Comitê de Ética em Pesquisa (CAAE: 56059822.2.0000.8052).

A clínica frequentada pelos participantes era de responsabilidade da pesquisadora e, portanto, foram tomados os devidos cuidados para que os responsáveis estivessem cientes de que a participação na pesquisa não acarretaria nenhum prejuízo ao tratamento dos participantes. A coleta foi realizada em horários combinados com os responsáveis de modo que não tivesse sua rotina habitual (e.g., escola e terapias) prejudicada. Nos dias em que as terapias iniciavam no final do turno, a coleta ocorria antes; já quando as terapias iniciavam no início do turno, a coleta ocorria depois delas.

Das seis crianças inicialmente recrutadas, três foram excluídas da pesquisa por: desistência do responsável, não ecoar as letras na elaboração dos vídeos e ter entrado de férias. Seus dados não serão apresentados. As crianças cujos resultados serão apresentados foram: o Participante 1, um menino de quatro anos, Nível 1 e começando a preencher habilidades presentes no Nível 2 no VB-MAPP (em avaliação realizada um mês antes da coleta), recebendo 20 hr semanais de aplicação ABA há um ano e um mês; a Participante 2, uma menina de quatro anos, Nível 1 e começando a preencher habilidades presentes no Nível 2 no VB-MAPP (em avaliação realizada um mês antes da coleta), recebendo 30 hr semanais de aplicação ABA há dois anos; e o Participante 3, um menino de seis anos, que estava começando a preencher habilidades do Nível 1 no VB-MAPP (em avaliação realizada dois meses antes da coleta), recebendo 20 hr semanais de aplicação ABA há cinco meses. Todos os participantes frequentavam a escola regularmente.

Ambiente, Equipamentos e Materiais

A pesquisa ocorreu na casa do participante e na clínica, controlando o ambiente da coleta de modo a eliminar ao máximo possível quaisquer distratores. Quando realizada na casa do participante, estiveram presentes os pais da criança, e quando no espaço da clínica estiveram presentes apenas a pesquisadora e uma assistente de pesquisa. Em ambos os ambientes, foi necessário o uso de um computador, celular, *tablet* ou outro dispositivo onde foram exibidos os vídeos para cada participante: um vídeo de par e um vídeo do próprio participante. Também foram utilizados: 21 cartões, possuindo 7 cm de largura e 7 cm de comprimento, com consoantes impressas na cor preta e fonte Arial tamanho 90; folhas de registro; reforçadores potenciais identificados em uma avaliação de preferência; e uma lista de checagem a ser preenchida pelos responsáveis indicando a apresentação dos vídeos para as crianças participantes.

Delineamento Experimental

Seguindo o delineamento de Marcus e Wilder (2009), foi utilizado um delineamento misto de linha de base múltipla entre participantes e de tratamentos alternados, com as fases: linha de base (LB), sonda pós-elaboração de vídeo (SV), e videomodelação (VM) — esta última envolvendo a apresentação alternada das condições de videomodelação por si mesmo e por par. A LB foi precedida por uma avaliação de preferência para escolha de itens potencialmente reforçadores; antes da SV houve a elaboração dos vídeos; e a VM foi precedida por uma avaliação de preferência por vídeos. Todas as fases são descritas detalhadamente a seguir.

Variável Dependente e Independente

A variável dependente do estudo foi a resposta vocal emitida pelo participante (correspondente ou não à consoante apresentada no cartão pela pesquisadora) e a variável independente foi a videomodelação por si ou por par.

Procedimento

Avaliação de Preferência para Escolha de Itens Potencialmente Reforçadores

Antes de iniciar a LB, foi realizada uma avaliação de preferência de estímulos múltiplos sem reposição (DeLeon & Iwata, 1996), para identificar os itens reforçadores potenciais a serem utilizados nas condições seguintes. Os objetos apresentados na avaliação foram cinco dentre 10 itens de interesse do participante recomendados pelo responsável. A avaliação consistiu em apresentar cinco itens ao participante de uma vez só e aguardar sua escolha. Após uma escolha, o item escolhido foi retirado do alcance do participante e os itens restantes foram reorganizados e apresentados novamente. Caso o participante não escolhesse nenhum dos itens apresentados na avaliação, foram apresentados os outros cinco itens de interesse, conforme a sugestão dos responsáveis, e o mesmo procedimento foi realizado com o novo conjunto de cinco itens. Os estímulos foram apresentados apenas

uma vez, e este procedimento foi realizado sucessivamente até que todos os cinco itens fossem escolhidos. Os itens utilizados como reforçadores potenciais foram os três primeiros escolhidos: quebra-cabeça, balão e massinha de modelar para o Participante 1; livros, turma da Mônica e Pocoyo para o Participante 2; e celular, quebra-cabeça e *slime* para o Participante 3.

Linha de Base

Nessa fase, todas as consoantes do alfabeto português foram apresentadas para cada participante, em ordem randômica, em 21 tentativas, ou seja, uma consoante para cada tentativa da sessão, totalizando 21 tentativas por sessão. A cada tentativa, a pesquisadora apresentou um dos 21 cartões com as consoantes do alfabeto português e perguntou ao participante: “Que letra é essa?”. Em cada tentativa, quando o participante respondia corretamente, um objeto de preferência (i.e., reforçador potencial) foi apresentado por 30 s, foi fornecido reforço social e a resposta foi registrada como correta; quando respondia incorretamente, uma nova tentativa era iniciada e a resposta era registrada como incorreta. Não houve apresentação de dicas na LB (nem em nenhuma outra condição), visto que esse procedimento poderia facilitar a instalação das respostas textuais já na LB, levando à exclusão de participantes. Após cada três tentativas foi solicitado ao participante alguma demanda previamente aprendida, de baixo custo de resposta e não relacionada ao comportamento textual—a qual foi consultada no prontuário da criança, a fim de que ela tivesse acesso a um item preferido, aumentasse a probabilidade de se manter engajada na atividade e não ficasse exposta a erros consecutivos diante de habilidades não aprendidas nesta condição. A quantidade de sessões de LB foi diferente para cada participante, a depender da estabilidade das respostas do próprio participante e do participante anterior. Assim, um participante só iniciava a condição de videomodelação em um momento diferente da finalização da LB do participante anterior.

Considerando que os participantes tiveram quantidades de erros diferentes em relação às 21 consoantes apresentadas, foi utilizada uma padronização para a seleção das 10 consoantes utilizadas nas fases seguintes, sendo que metade foi designada aleatoriamente para a videomodelação por si e metade para a videomodelação por pares. Quando o participante acertou a consoante em três sessões consecutivas ela não foi utilizada na fase de videomodelação; quando o participante errou em três sessões consecutivas ela foi elegível para fase de videomodelação; porém, quando o participante não errou nas três sessões o número suficiente de consoantes, foram utilizadas na fase de videomodelação consoantes com dois erros consecutivos. Se o participante apresentasse erros em três sessões em mais de 10 consoantes, a escolha das consoantes foi feita de forma randômica. A Tabela 1 apresenta as consoantes selecionadas para cada participante.

Tabela 1

Consoantes Seleccionadas Para as Condições de Videomodelação por Si Mesmo e por Par de Cada Participante

Participantes	Consoantes para Vídeo de Si	Consoantes para Vídeo de Par
Participante 1	D, K, L, Q, T	F, H, J, N, R
Participante 2	B, C, H, Q, Z	G, M, S, T, V
Participante 3	B, C, G, M, T	F, H, P, V, Z

Elaboração dos Vídeos

Para a produção do vídeo de si mesmo, inicialmente a pesquisadora filmou o próprio participante, sendo fornecido um modelo vocal da resposta para o participante ecoar, sem qualquer contato com a letra escrita. Imediatamente depois que o participante ecoasse corretamente, ele teve acesso a um item reforçador potencial por 30 s. Em outro momento, a pesquisadora se filmou apresentando o cartão com as letras escolhidas e perguntando “que letra é essa?” Esses vídeos foram editados para que, quando apresentados para o participante, ele se visse no vídeo emitindo a resposta vocal. Foram gravados vídeos de si com as cinco consoantes seleccionadas a partir dos erros do participante na LB. As outras cinco consoantes foram utilizadas para a gravação dos vídeos de par, sendo filmados com um colaborador. As crianças utilizadas como pares na pesquisa apresentaram características físicas, altura e cor de pele similares, além de terem o mesmo sexo e idade dos participantes para os quais foram par. A produção dos vídeos de par seguiu o mesmo procedimento dos vídeos de si mesmo, com a alteração apenas do modelo empregado. Também foram gravados vídeos de si e de par com outras cinco consoantes não seleccionadas para a fase VM a fim de utilizá-los na avaliação de preferência dos vídeos. Não se utilizou as mesmas consoantes empregadas na fase VM para reduzir a exposição dos participantes a elas. A gravação de cada vídeo de si necessitou de cerca de 10 min, enquanto a gravação de cada vídeo de par, cerca de 6 min.

Sonda Pós Elaboração de Vídeos

Após a elaboração dos vídeos, cada participante passou por uma única sessão de LB adicional, para checar se houve aprendizagem durante a preparação dos vídeos. Essa sonda foi idêntica às sessões de LB anteriores, com exceção que aqui foram testadas apenas as 10 consoantes utilizadas na elaboração dos vídeos.

Avaliação de Preferência por Vídeos

Essa fase identificou a preferência dos participantes em relação aos dois tipos de vídeos (de par e de si mesmo). Foi utilizada a avaliação de preferência de escolha pareada (Silveira et al., 2017). Nos vídeos dessa fase foram utilizadas consoantes diferentes das selecionadas para a fase de VM tanto nos vídeos de par quanto de si mesmo. O procedimento foi realizado ao longo de três sessões. Em todas as três sessões foram apresentados os mesmos vídeos. Todas as sessões foram idênticas e ocorreram da seguinte forma: a pesquisadora apresentou em um celular um vídeo de par e em outro celular exatamente do mesmo modelo um vídeo de si mesmo, lado a lado. O participante teve acesso a 10 s de cada vídeo, com a ordem de apresentação dos vídeos sendo previamente definida de forma randomizada. Após isso, a pesquisadora solicitou que o participante escolhesse um dos vídeos para assistir. Após a escolha, o vídeo era iniciado e a escolha foi registrada na folha de registro. Após as três sessões, uma em cada dia, o vídeo mais escolhido foi considerado o preferido.

Videomodelação

Esta fase foi composta por três etapas. Na primeira etapa, os responsáveis por cada participante foram instruídos a apresentar o vídeo para a videomodelação por par ou por si mesmo, aquele que houvesse sido previamente sorteado para a etapa, em dois dias consecutivos, três vezes ao dia: antes da escola, imediatamente depois da escola, e à noite, antes de dormir. Os responsáveis pelos participantes receberam um cronograma com horários nos quais os vídeos deveriam ser exibidos. A pesquisadora entrou em contato com os responsáveis para lembrá-los de mostrar os vídeos no horário combinado, anotar o ocorrido durante a exibição e realizar uma chamada de vídeo para observar se existiam distratores no ambiente. Foi disponibilizado ainda uma lista de checagem para ser preenchida pelos responsáveis durante os dias de videomodelação que aconteceram na casa do participante.

Após os dois dias, os participantes compareceram à clínica, onde a pesquisadora exibiu o vídeo apresentado em casa e realizou uma tentativa com cada uma das cinco consoantes. Como nas tentativas da LB, cada tentativa iniciava quando a pesquisadora exibia o cartão com a consoante e perguntava ao participante: “Que letra é essa?”. O participante teve 10 s para responder; caso não respondesse (i.e., ausência de resposta) ou respondesse de modo incorreto, a pesquisadora registrava o erro e seguia para a outra tentativa. Caso a criança respondesse corretamente, recebia reforço social e tinha acesso por 30 s a um item preferido e o acerto era registrado. A cada três tentativas, também como na LB, foi solicitado ao participante alguma demanda pré-definida de baixo custo de resposta e não relacionada a comportamento textual.

Na segunda etapa, os responsáveis apresentaram o segundo tipo de vídeo—aquele que o participante não assistiu. Assim, se na primeira etapa o participante assistiu ao vídeo de si mesmo, agora assistia ao vídeo de par ou vice-versa. Depois dos dois dias consecutivos de treino em casa, a pesquisadora fez o teste da segunda

etapa de videomodelação. Dessa forma, nas duas primeiras etapas da fase de videomodelação, os responsáveis pelos participantes apresentavam um dos dois vídeos, três vezes ao dia, em dois dias seguidos. Nesses momentos, entretanto, não havia a apresentação de tentativas de ensino (i.e., qualquer forma de reforço diferencial programado) nem mesmo qualquer forma de registro das respostas dos participantes. Assim, nenhuma sessão experimental foi realizada na casa dos participantes; essas atividades realizadas pelos responsáveis não compõem propriamente o delineamento de tratamento alternado realizado, consistindo em pré-exposições às tentativas de videomodelação realizadas na clínica. Ao final da primeira e da segunda etapa, ocorreram na clínica uma sessão experimental para cada vídeo – essas sim, compondo o delineamento experimental.

Já na terceira etapa, os dois tipos de vídeo foram apresentados para cada participante apenas pela pesquisadora na clínica, sem sua apresentação prévia pelos responsáveis em casa, com a ordem de apresentação deles sendo alternada a cada dia. Por exemplo, se o primeiro dia iniciasse com o vídeo de par, o segundo dia iniciaria com o vídeo de si próprio e assim sucessivamente. Logo depois da exibição de cada vídeo, a pesquisadora realizou cinco tentativas, cada uma referente a uma consoante exibida no vídeo apresentado. Cada conjunto de cinco tentativas foi considerado uma sessão. O critério de estabilidade para o encerramento dessa condição foi o mesmo utilizado por Marcus e Wilder (2009), isto é, pelo menos 80% de respostas corretas por três sessões consecutivas em alguma das condições de VM; sendo necessário, no mínimo, o acerto de quatro das cinco tentativas por três sessões consecutivas em uma das duas condições de VM. Se o participante atingisse o critério em uma condição, mas não na outra, seriam realizadas mais cinco sessões para que ele pudesse atingir o critério em ambas as condições. Caso a estabilidade não fosse atingida em nenhuma das condições, essa etapa seria encerrada com o máximo de 30 sessões.

Avaliação de Integridade da Aplicação e Concordância entre Observadores

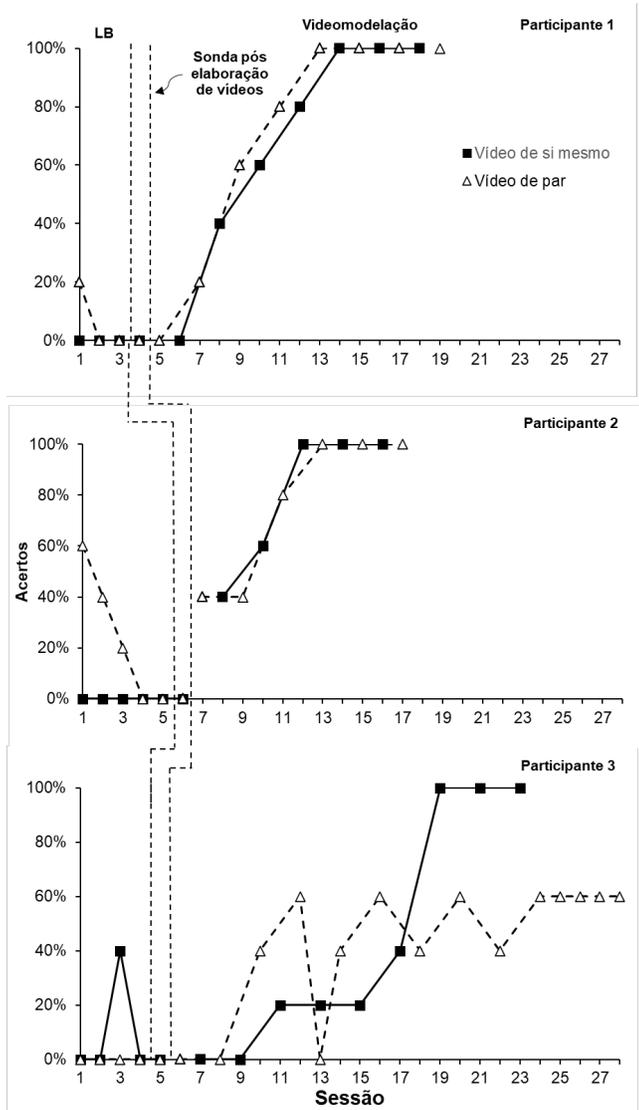
Os dados foram registrados em uma folha de registro pela pesquisadora e por uma assistente de pesquisa que ficou em um canto da sala de coleta. A assistente também ficou responsável em registrar a integridade da aplicação da pesquisadora, a qual foi de 100% para todos os participantes em todas as condições, com exceção da terceira sessão da LB do Participante 1 que apresentou 95%.

A concordância entre observadores foi calculada através da divisão do número de concordâncias multiplicado por 100 pela soma das concordâncias e das discordâncias. O cálculo de concordância foi realizado ao final de cada sessão. A concordância em todas as fases do estudo para todos os participantes foi sempre igual ou acima de 80%, com a média dentre todas as fases igual a 98% para o Participante 1, 99% para o Participante 2 e 95% para o Participante 3.

Resultados e Discussão

A Figura 1 apresenta a porcentagem de tentativas com respostas corretas em todas as condições para os três participantes. Na LB são apresentados os dados referentes apenas às consoantes que foram usadas posteriormente na fase de VM. Ainda que na LB os participantes tenham apresentado alguns acertos, as consoantes foram elegíveis para a fase de VM pois apresentaram dois ou mais erros consecutivos. Os acertos na LB das consoantes selecionadas para a videomodelação de par da Participante 2 diminuíram ao longo das sessões, ainda que as respostas corretas tenham sido reforçadas: na primeira sessão, houve acertos em M, S e V; na segunda sessão, em M e V; e na terceira sessão, somente um acerto na consoante M. Apesar das respostas textuais às consoantes M e V terem sido reforçadas mais vezes, não atingiram critério para serem excluídas da pesquisa. Na sonda pós-elaboração do vídeo, entretanto, não houve respostas corretas das consoantes selecionadas para a fase de VM.

Figura 1
Porcentagem de Respostas Corretas dos Participantes em Todas as Condições Experimentais Para as Consoantes Empregadas no Vídeo de Si e no Vídeo de Par



Todos os participantes aumentaram a emissão de respostas corretas durante a fase de VM e atingiram o critério de encerramento tanto na condição de vídeo de si mesmo quanto na de vídeo de par, com exceção do Participante 3 na condição de vídeo de par. O Participante 1 atingiu 100% de acertos em ambas as condições de videomodelação após cinco sessões, mantendo essa porcentagem por três sessões na condição com vídeos de si mesmo e por quatro sessões na condição com vídeos de par¹. Na videomodelação de si mesmo, a Participante 2 atingiu 100% de acertos já após três sessões, enquanto na videomodelação de par, após quatro sessões. O Participante 3 atingiu o critério de estabilidade com nove sessões de videomodelação na condição de vídeo de si mesmo. Na videomodelação por par, ele não atingiu o critério e a condição foi encerrada cinco sessões após ele ter atingido critério no vídeo de si. O maior número de sessões para atingir o critério na condição de vídeo de si e a estabilidade com apenas 50% de acertos na condição de vídeo de par do Participante 3 podem ter se devido ao repertório inicial desse participante, o único em intervenção por menos de um ano (apenas cinco meses), também o único que ainda estava atingindo habilidades do Nível 1 de ajuda—enquanto os outros participantes já estavam atingindo habilidades de Nível 2 no VBMAPP. Além disso, durante a elaboração dos vídeos, os participantes ecoaram somente as letras do vídeo de si, implicando em uma maior exposição às consoantes utilizadas nessa condição, o que pode ter contribuído para o melhor desempenho na condição de vídeo de si do Participante 3. Esse aspecto do procedimento da presente pesquisa pode ser apontado como uma limitação, sendo interessante, portanto, que em estudo futuros o modelo ecoico dos alvos de ambas as condições seja fornecido para que os participantes sejam expostos aos estímulos igualmente.

Os resultados de Marcus e Wilder (2009) apontaram para a videomodelação por si como superior à videomodelação por par para o ensino de respostas textuais, sugerindo que as diferenças físicas dos pares com os participantes possam ter influenciado nesse resultado. Diferente destes resultados, nesta pesquisa, os Participantes 1 e 2 apresentaram quantidade de acertos semelhantes em ambas as condições de videomodelação, resultado semelhante ao de Sherer et al. (2001). Entretanto, o fato do Participante 2 ter apresentado uma quantidade maior de acertos na videomodelação de si (100%) comparado a de par (80%) durante a primeira sessão em que atinge critério, e o fato do Participante 3 não ter atingido o critério de estabilidade na condição de videomodelação de par, corrobora com os dados de Marcus e Wilder (2009). De qualquer forma, se houve alguma diferença entre as modalidades de videomodelação, elas não foram tão marcadas quanto as observadas por Marcus e Wilder—o que pode ter se devido à similaridade física dos pares em relação aos participantes utilizada na presente pesquisa.

Na Avaliação de Preferência de Vídeos, todos os participantes demonstraram preferência pelos vídeos de par, escolhendo duas vezes o vídeo de par e uma vez o vídeo de si mesmo, diferentemente das observações anedóticas de Marcus e Wilder (2009) as quais sugeriram que seus participantes preferiram mais os vídeos de si. Esse resultado pode se dar pelos pares e participantes desta pesquisa apresentarem semelhanças físicas na altura e na cor e terem o mesmo sexo e idade, enquanto Marcus e Wilder (2009) apontam que seus pares não eram semelhantes aos seus partici-

pantes. Bandura et al. (1961) apontaram para uma maior probabilidade de indivíduos imitarem pares quando estes possuem características semelhantes à dos indivíduos, fato que também pode explicar as semelhanças nos desempenhos entre as condições de par e de si dos Participantes 1 e 2. Futuras pesquisas podem comparar os desempenhos na videomodelação com pares semelhantes e não semelhantes, além de avaliar a preferência das crianças a assistir vídeos de diferentes pares.

Os participantes desta pesquisa terem atingido os critérios de estabilidade em menos sessões do que Marcus e Wilder (2009) pode ser explicado pelas letras utilizadas serem consoantes presentes no alfabeto português, língua materna de todos os participantes. Portanto, estes podem ter recebido, mesmo que inadvertidamente, ensino relacionado às consoantes em suas histórias pré-experimentais, em contraste com o procedimento de Marcus e Wilder, que utilizou letras do alfabeto grego.

Os resultados desta pesquisa corroboram pesquisas anteriores demonstrando que crianças com TEA podem ser beneficiadas com um ensino rápido e com baixo custo de respostas textuais por videomodelação. Os resultados apontaram para uma emissão de respostas textuais levemente mais precisa (ver Figura 1) durante a condição de videomodelação por si mesmo. A utilização de pares semelhantes pode ter sido uma influência positiva no desempenho durante a videomodelação de par, considerando que indivíduos possuem maiores chances de imitar quando seus pares possuem semelhanças físicas (Bandura et al., 1961). A pesquisa utilizou como alvo de ensino as consoantes do alfabeto português, língua materna dos participantes, por se tratar de estímulos socialmente relevantes. Os profissionais da área devem se atentar para o fato de que não são todas as crianças com TEA que podem ser expostas ao ensino por videomodelação, sendo preciso alguns requisitos, como as habilidades de manter-se atento ao vídeo, imitação generalizada e imitação atrasada (MacDonald et al., 2015).

A decisão sobre qual tipo de videomodelação empregar no ensino deve levar em conta que a videomodelação por si mesmo limita as vantagens advindas da padronização do procedimento e da facilidade na produção dos materiais de ensino. Essas duas vantagens se mantêm para o ensino da própria pessoa gravada em vídeo, mas não permite que o mesmo vídeo seja empregado em videomodelação por si mesmo com outras pessoas. Algo semelhante ocorre com a videomodelação por pares semelhantes, já que seria necessário encontrar pares semelhantes a cada aprendiz para gravar novos vídeos. Além disso, a videomodelação por si mesmo possivelmente não se aplica ao ensino de qualquer classe de resposta. No presente estudo, o ensino de respostas textuais de consoantes foi possível devido à apresentação do estímulo discriminativo (i.e., pesquisadora perguntar qual a letra e mostrá-la) e da resposta modelo (i.e., a própria criança vocalizar a consoante) poderem ser gravadas separadamente e unidas na edição. Além disso, a resposta modelo pôde ser evocada por meio do repertório de ecóico já instalado nos participantes. Entretanto, no caso de respostas mais complexas, envolvendo por exemplo cadeias de respostas (e.g., vestir uma roupa) ou interações sociais mais prolongadas (e.g., brincar com pares), criar um vídeo equivalente seria muito mais custoso ou impossível.

Apesar da literatura sobre videomodelação para pessoas com TEA mostrar efeitos positivos para ensino de respostas textuais, ainda é necessário que em pesquisas

futuras possa ser investigada a comparação da videomodelação de si mesmo e de par para o ensino de respostas textuais sob controle de estímulos mais complexos, como palavras e frases. A pesquisa atual sugere ainda estudos que continuem a avaliar esse tipo de procedimento para o ensino de outras habilidades dentro do processo de leitura e escrita, como transcrição, busquem avaliar a preferência por pares entre os participantes, ofereçam modelo ecoico dos alvos de ambas as condições para que os participantes sejam expostos aos estímulos de forma igual, comparem os desempenhos na videomodelação com pares semelhantes versus não semelhantes e avaliem a preferência das crianças a assistir vídeos de diferentes pares.

Referências

- Acar, G., Altun, G. P., Yurdalan, S., & Polat, M. G. (2016). Efficacy of neurodevelopmental treatment combined with the Nintendo® Wii in patients with cerebral palsy. *Journal of Physical Therapy Science*, 28(3), 774-780. <https://doi.org/10.1589/jpts.28.774>
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th Ed.). <https://doi.org/10.1176/appi.books.9780890425787>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1(1), 91-97. <http://doi.org/10.1901/jaba.1968.1-91>
- Bandura, A., Ross, D., & Ross, S. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 63, 575-582. <https://doi.org/10.1037/h0045925>
- Barboza, A. A., Silva, A. J. M., Barros, R. S., & Higbee, T. S. (2015). Efeitos de videomodelação instrucional sobre o desempenho de cuidadores na aplicação de programas de ensino a crianças diagnosticadas com autismo. *Acta Comportamental*, 23(4), 405-421. <http://www.redalyc.org/articulo.oa?id=274543456004>
- Barros, R. S. (2003). Uma introdução ao comportamento verbal. *Revista Brasileira de Terapia Comportamental e Cognitiva*, 5(1), 73-82. <https://doi.org/10.31505/rbtcc.v5i1.92>
- Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children*, 73, 264-287. <https://doi.org/10.1177/001440290707300301>
- Brasilense, I. C. S., Flores, E. P., Barros, R. S., & Souza, C. B. A. (2018). Aprendizagem observacional em crianças com autismo: Efeitos do ensino de respostas de monitoramento via videomodelação. *Psicologia: Teoria e Pesquisa*, 34, e3424. <https://doi.org/10.1590/0102.3772e3424>
- Catania, A. C. (1999). *Aprendizagem: Comportamento, linguagem e cognição*. Artmed.
- Charlop, M., Lang, R., & Rispoli, M. (2018). *Play and social skills for children with autism spectrum disorder*. Springer Cham. <https://doi.org/10.1007/978-3-319-72500-0>

- Decker, M. M., & Buggey, T. (2014). Using video self- and peer modeling to facilitate reading fluency in children with learning disabilities. *Journal of Learning Disabilities, 47*(2), 167-177. <https://doi.org/10.1177/0022219412450618>
- DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*(4), 519-533. <https://dx.doi.org/10.1901%2Fjaba.1996.29-519>
- Guerra, B. T., & Almeida-Verdu, A. C. M. (2016). Ensino de operantes verbais em pessoas com Transtorno do Espectro Autista no The Analysis of Verbal Behavior: Revisão sistemática. *Revista Brasileira de Terapia Comportamental e Cognitiva, 18*(2), 73-85. <https://doi.org/10.31505/rbtcc.v18i2.884>
- MacDonald, R. P. F., Dickson, C. A., Martineau, M., & Ahearn, W. H. (2015). Prerequisite skills that support learning through video modeling. *Education and Treatment of Children, 38*(1), 33-47. <https://doi.org/10.1353/etc.2015.0004>
- Marcus, A., & Wilder, D. A. (2009). A comparison of peer video modeling and self video modeling to teach textual responses in children with autism. *Journal of Applied Behavior Analysis, 42*(2), 335-341. <https://doi.org/10.1901/jaba.2009.42-335>
- Martone, M. C. C., & Santos-Carvalho, L. H. Z. (2012). Uma revisão dos artigos publicados no Journal of Applied Behavior Analysis (JABA) sobre comportamento verbal e autismo entre 2008 e 2012. *Perspectivas em Análise do Comportamento, 3*(2), 73-86. <https://doi.org/10.18761/perspectivas.v3i2.227>
- Organização Pan-Americana da Saúde; Organização Mundial da Saúde – OPAS/OMS (2023). *Transtorno do Espectro Autista*. <https://www.paho.org/pt/topicos/transtorno-do-espectro-autista>
- Prater, M. A., Carter, N., Hitchcock, C., & Dowrick, P. (2012). Video self-modeling to improve academic performance: A literature review. *Psychology in the Schools, 49*(1), 71-81. <https://doi.org/10.1002/pits.20617>
- Sherer, M., Pierce, K. I., Paredes, S., Kisacky, K. L., Ingersoll, B., & Schreibman, L. (2001). Enhancing conversation skills in children with autism via video technology: Which is better, “self” or “other” as a model? *Behavior Modification, 25*, 140-158. <https://doi.org/10.1177/0145445501251008>
- Silveira, F., Panosso, M., Dal Ben, R., & Gallano, T. (2017). Métodos de avaliação de itens de preferência para a identificação de reforçadores. *Revista Brasileira de Terapia Comportamental e Cognitiva, 19*, 89-107. <https://doi.org/10.31505/rbtcc.v19i2.1034>
- Skinner, B. F. (1992). *Verbal Behavior*. B. F. Skinner Foundation. (Trabalho original publicado em 1957).
- Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., & Savage, M. N. (2020). *Evidence-based practices for children, youth, and young adults with autism*. FPG Child Development Institute. <https://eric.ed.gov/?id=ED609029>
- Sundberg, M. L. (2014). *VB-MAPP: Verbal Behavior Milestones Assessment and Placement Program. Guide* (2nd Ed.). AVB Press.
- Varella, A. A., & de Souza, C. M. C. (2018). Ensino por tentativas discretas: Revisão sistemática dos estudos sobre treinamento com vídeo modelação. *Revista*

Brasileira de Terapia Comportamental e Cognitiva, 20(3), 73-85. <https://doi.org/10.31505/rbtcc.v20i3.1215>

(Received: July 26, 2023; Accepted: October 20, 2023)

Notas

¹ Uma sessão a mais foi realizada na condição de videomodelação por par devido a um erro de procedimento.