

João Paulo Araújo Lessa



**MENSURAÇÃO E CORREÇÃO DE ERROS SISTEMÁTICOS
NA AVALIAÇÃO SOCIOEMOCIONAL: UM ESTUDO
PSICOMÉTRICO SOBRE VINHETAS-ÂNCORA**

Apoio:

Universidade São Francisco | Instituto Ayrton SENNA



**CAMPINAS
2022**

João Paulo Araújo Lessa

**MENSURAÇÃO E CORREÇÃO DE ERROS
SISTEMÁTICOS NA AVALIAÇÃO
SOCIOEMOCIONAL: UM ESTUDO
PSICOMÉTRICO SOBRE VINHETAS-ÂNCORA**

Tese apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco, Área de Concentração - Avaliação Psicológica, para obtenção do título de Doutor.

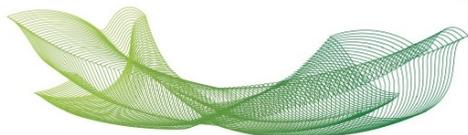
ORIENTADOR: PROF. DR. RICARDO PRIMI

CAMPINAS
2022

UNIVERSIDADE SÃO FRANCISCO

PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU *EM PSICOLOGIA*

MESTRADO/DOCTORADO



Educando
para a paz

PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM PSICOLOGIA

João Paulo Araújo Lessa defendeu a tese “**MENSURAÇÃO E CORREÇÃO DE ERROS SISTEMÁTICOS NA AVALIAÇÃO SOCIOEMOCIONAL: UM ESTUDO PSICOMÉTRICO SOBRE VINHETAS-ÂNCORA**” aprovado pelo Programa de Pós-Graduação Stricto Sensu em Psicologia da Universidade São Francisco em 29 de julho de 2022 pela Banca Examinadora constituída por:

Prof. Dr. Ricardo Primi
Orientador e Presidente

Prof. Dr. Víthor Rosa Franco
Examinador

Prof. Dr. Felipe Valentini
Examinador

Prof. Dr. Evandro Morais Peixoto
Examinador

Prof. Dr. Daniel Domingues dos Santos
Examinador

Prof. Dr. Alexandre José de Souza Peres
Examinador

Dedicatória

*A Mário e Iolita
(in memoriam)*

Agradecimentos

Aos meus pais, Ricardo Lessa e Seldja Araújo, bem como aos meus irmãos Ricardo Filho e Vinícius, pelo apoio em todo o percurso.

Aos colegas do PPG, que de tanto são, ficaria impossível colocar em uma página!

Aos companheiros de Laboratório 3, que tanto compartilhamos café, biscoitos e bolachas, discussões sobre os trabalhos mais diversos, e, além de tudo, risadas, memes e uma amizade sólida! Vocês deixaram essa caminhada mais acessível. Fernanda Otoni, Léo Mose, Pedro Cortez e Yara Padilha, obrigado pelas memórias!

Ao LabAPE, pelos ensinamentos, em especial A Karina Oliveira, sendo uma das pessoas mais competentes que pude conhecer nessa trajetória acadêmica.

A Jullie, por todo o companheirismo durante este processo.

Aos professores que também foram orientadores neste meu ciclo acadêmico. Acácia Angeli, Ana Paula Porto Noronha, Ana Elisa, Cristian Zanon, Lucas Francisco, Rodolfo Ambiel e Nelson Hauck por todo o ensinamento durante o percurso. Ainda, e em especial, Felipe Valentini, um professor que confiou no meu potencial e abriu diversas portas, além de estar sempre disposto a ensinar e tirar dúvidas.

Ao meu orientador (que também foi chefe por vários meses), Prof. Ricardo Primi, que não somente é reconhecido enquanto pesquisador exímio na Psicologia, mas professor que também soube me acolher, ensinar e cobrar, tendo me indicado para o doutorado direto, com uma bolsa de pesquisa, em um momento em que a Ciência no país estava começando a ser fortemente atacada, colaborando indescritivelmente no meu desenvolvimento enquanto pesquisador e profissional. A ele, que possivelmente viu algo em mim que eu nunca vi, gratidão!

Ao Instituto Ayrton SENNA (IAS), em especial à Gisele Alves, gerente executiva da vice-presidência de Evidências, pelo apoio financeiro nos 34 meses ofertados com a bolsa de pesquisa e à Universidade São Francisco pelo convênio estabelecido com o IAS, sem os quais este trabalho não teria acontecido.

Resumo

Lessa, J.P.A. (2022). *Mensuração e correção de erros sistemáticos na avaliação socioemocional: Um estudo psicométrico sobre Vinhetas-âncora*. Tese de Doutorado, Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Campinas.

A avaliação socioemocional tem sido considerada como um *hot topic*, abrangendo áreas como a Psicologia, Pedagogia, Economia, dentre outras. A forma comumente utilizada para se realizar esta avaliação é por meio de instrumentos de autorrelato, em que o avaliando é requerido a ler itens que avaliam determinadas características e, em uma escala Likert, ele endossa uma das opções apresentadas. Ainda que seja um método bastante difundido, o autorrelato é suscetível a erros do tipo sistemático, sendo os estilos de resposta uma das principais fontes destes erros. As pesquisas sobre o tema são inúmeras, já que o impacto que os estilos de respostas trazem para os instrumentos afetam diretamente as evidências de validade e estimativas de precisão dele, assim como a equidade e justiça da avaliação. Neste sentido, estudos direcionados a investigar diferentes técnicas e métodos de controle de erros sistemáticos são necessários para compreender ainda mais o funcionamento dos estilos de resposta e seus impactos nas propriedades psicométricas de instrumentos de autorrelato. Dentre as técnicas propostas para se realizar um controle destas fontes de erro está a técnica de Vinhetas-âncora, que surge no âmbito das ciências políticas para controlar o que os pesquisadores perceberam como respostas incongruentes. Esta técnica utiliza situações hipotéticas acerca do construto que se pretende avaliar e requer dos avaliandos que respondam a estas situações, configurando um método de heterorrelato. Diante do exposto, o objetivo da presente tese foi investigar a validade das Vinhetas-âncora como forma de medir os estilos de resposta. Para alcançar este objetivo geral, quatro objetivos específicos foram elaborados, sendo um estudo para cada. O primeiro estudo buscou compreender como as escalas de respostas às Vinhetas-âncora são utilizadas. O segundo objetivou compreender qual a relação do desempenho das Vinhetas-âncora com variáveis sociodemográficas, de desempenho escolar e competências socioemocionais. Por sua vez, o terceiro averiguou a possibilidade de mensurar o estilo de resposta aquiescente por meio das Vinhetas-âncora e sua relação com outras medidas de aquiescência. Por fim, o quarto estudo investigou como as Vinhetas-âncora podem medir o viés de referência de grupo e sua relação com demais medidas deste estilo de resposta. Os resultados do primeiro estudo demonstraram que a utilização das Vinhetas-âncora é mais consistente com situações em que o traço avaliado é positivo. No segundo estudo, os resultados demonstraram que as meninas apresentaram maior relação com o uso consistente das Vinhetas-âncora, e de forma semelhante esta relação foi encontrada com melhores notas nos desempenhos de Linguagens e Matemática. O terceiro estudo apresentou que as correlações entre os índices de mensuração de aquiescência criados para as Vinhetas-âncora foram estatisticamente significativas com índices de aquiescência externos. Por fim, o quarto estudo demonstrou uma relação significativa entre a dispersão do uso da escala com variáveis externas. Tais resultados sustentam a hipótese de que a técnica de Vinhetas-âncora pode ser uma solução para mensurar estilos de resposta aquiescente e de viés de grupo, avançando no tema ao permitir uma melhor compreensão de como o respondente utiliza a escala de resposta.

Palavras-chave: avaliação educacional, avaliação da personalidade, autorrelato, estilos de respostas.

Abstract

Lessa, J.P.A. (2022). *Measurement and correction of systematic errors in socioemotional assessment: A psychometric study about Anchoring vignettes*. Master's Thesis, Post-Graduate Studies in Psychology, University San Francisco, Campinas, São Paulo.

Socio-emotional assessment has been considered a hot topic, covering areas such as Psychology, Pedagogy, and Economics, among others. Commonly, this kind of assessment is conducted using self-report instruments, in which the subject is required to read items assessing specific characteristics and, on a Likert scale, endorse one of the options presented. Self-report, while widespread, is susceptible to systematic errors, with response styles being one of the main sources of these errors. Numerous studies have been conducted on the topic, as styles of responses directly contribute to evidence of validity, accuracy estimates, as well as equity and fairness of the assessment. Studies aimed at investigating different techniques and methods are needed to better understand how response styles' function and how they affect self-report instruments' psychometric properties. A technique that appears in the field of political science to control what researchers perceive as incongruous responses is the Anchor-Vignette technique. Hypothetical situations are posed regarding the construct to be evaluated and the subjects are required to respond in response, thereby configuring a heteroreporting method. Due to the foregoing, the present study investigated the validity of Anchor Vignettes as a means of assessing response styles. To reach this general objective, four specific objectives were elaborated, being one study for each one. As part of the first study, we aimed to understand the scales of responses to Anchor Vignettes. In the second study, we examined the relationship between Anchor Vignette's performance and sociodemographic factors, school performance and socioemotional skills. A third study investigated the possibility of measuring acquiescent response style using Anchor Vignettes and its relationship with other measures of acquiescence. In the fourth study, we investigated how Anchor Vignettes can measure group reference bias and how they relate to other measures of this response style. According to the first study, anchor vignettes are more appropriate when the trait being assessed is positive. As in the first study, there was a stronger correlation between Anchor Vignettes and language and math performance in the second study. In the third study, correlations between the acquiescence measurement indices created for the Anchor Vignettes and external acquiescence indices were statistically significant. Finally, the fourth study demonstrated a significant relationship between the dispersion of use of the scale with external variables. Results of these experiments support the hypothesis that the Anchor Vignettes can be an effective method for measuring acquiescent response styles and group bias, allowing a deeper understanding of how the respondent uses the response scale. In this way, future studies will be able to evaluate different groups with different constructs.

Keywords: educational assessment, personality assessment, response style, self-report

Resumen

Lessa, J.P.A. (2022). *Medición y corrección de errores sistemáticos en la evaluación socioemocional: Un estudio psicométrico sobre Anchoring vignettes*. Tesis Doctoral, Programa de Estudios de Posgrado en Psicología, Universidad San Francisco, Campinas, São Paulo.

La evaluación socioemocional ha sido considerada un tema candente, abarcando áreas como la Psicología, la Pedagogía, la Economía, entre otras. La forma comúnmente utilizada para llevar a cabo esta evaluación es a través de instrumentos de autoinforme, en los que se requiere que el sujeto lea ítems que evalúen determinadas características y, en una escala tipo Likert, avale una de las opciones presentadas. Aunque es un método muy extendido, el autoinforme es susceptible de errores sistemáticos, siendo los estilos de respuesta una de las principales fuentes de estos errores. La investigación sobre el tema es numerosa, ya que el impacto que los estilos de respuesta traen a los instrumentos afectan directamente la evidencia de validez y las estimaciones de precisión del instrumento, así como la equidad y equidad de la evaluación. En este sentido, se necesitan estudios dirigidos a investigar diferentes técnicas y métodos para comprender mejor el funcionamiento de los estilos de respuesta y su impacto en las propiedades psicométricas de los instrumentos de autoinforme. Entre las técnicas propuestas para controlar estas fuentes de error se encuentra la técnica Anchor-Vignette, que aparece en el campo de la ciencia política para controlar lo que los investigadores perciben como respuestas incongruentes. Esta técnica utiliza situaciones hipotéticas sobre el constructo a evaluar y requiere que los sujetos respondan a estas situaciones, configurando un método de heteroinforme. Dado lo anterior, el objetivo del presente estudio fue investigar la validez de Anchor Vignettes como una forma de medir los estilos de respuesta. Para alcanzar este objetivo general se elaboraron cuatro objetivos específicos, siendo un estudio para cada uno. El primer estudio buscó comprender cómo se utilizan las escalas de respuestas a las Viñetas de Anclaje. El segundo tuvo como objetivo comprender la relación entre el desempeño de las Viñetas Anchor y variables sociodemográficas, desempeño escolar y habilidades socioemocionales. Por su parte, el tercero investigó la posibilidad de medir el estilo de respuesta aquiescente a través de las Viñetas Anchor y su relación con otras medidas de aquiescencia. Finalmente, el cuarto estudio investigó cómo Anchor Vignettes puede medir el sesgo de referencia grupal y su relación con otras medidas de este estilo de respuesta. Los resultados del primer estudio mostraron que el uso de Anchor Vignettes es más consistente con situaciones en las que el rasgo evaluado es positivo. En el segundo estudio, los resultados mostraron que las niñas tenían una mayor relación con el uso consistente de Anchor Vignettes, y de igual manera esta relación se encontró con mejores puntajes en lenguaje y desempeño en Matemáticas. El tercer estudio mostró que las correlaciones entre los índices de medición de aquiescencia creados para las viñetas de anclaje fueron estadísticamente significativas con los índices de aquiescencia externos. Finalmente, el cuarto estudio demostró una relación significativa entre la dispersión del uso de la escala con variables externas. Estos resultados apoyan la hipótesis de que la técnica Anchor Vignettes puede ser una solución para medir los estilos de respuesta aquiescentes y el sesgo grupal, avanzando en el tema al permitir una mejor comprensión de cómo el encuestado usa la escala de respuesta. De esta forma, se espera que futuros estudios puedan evaluar diferentes grupos con otros constructos.

Palabras clave: evaluación educativa, evaluación de la personalidad, autoinforme, estilos de respuesta

Sumário

LISTA DE FIGURAS	viii
LISTA DE TABELAS	ix
LISTA DE ABREVIATURAS E SÍMBOLOS	xi
APRESENTAÇÃO.....	1
INTRODUÇÃO.....	3
MÉTODO.....	30
PARTICIPANTES	30
INSTRUMENTOS	31
PROCEDIMENTOS	34
ANÁLISE DE DADOS	34
RESULTADOS	37
DISCUSSÃO.....	63
CONSIDERAÇÕES FINAIS	71
REFERÊNCIAS.....	72

Lista de figuras

Figura 1- Regras para correção não paramétrica das Vinhetas-âncora	18
Figura 2- Representação hipotética de endosso das Vinhetas-âncora	35
Figura 3- Histograma com as distribuições das respostas as Vinhetas-âncora	38
Figura 4- Histograma com as distribuições das médias de frequência das respostas nas Vinhetas-âncora....	40
Figura 5- Histograma das médias de aquiescência	47
Figura 6- Visualização da relação entre nota de Linguagens com os índices de aquiescência	51
Figura 7.- . Visualização da relação entre os índices de aquiescência no traço e no atributo com a nota em Linguagens	52
Figura 8- Histogramas com as frequências dos índices de viés de grupo	55
Figura 9- Visualizações gráficas da dispersão entre as notas de Linguagens e os índices gerais das médias e desvios-padrão de viés de grupo	58
Visualização gráfica da dispersão entre as notas em Linguagens com os índices de média de viés de grupo para a dimensão Engajamento com os outros	
Figura 11 - Visualização gráfica da dispersão entre os escores da dimensão Amabilidade com os índices gerais de média de viés de grupo.	62

Lista de tabelas

Tabela 1- Exemplo de três Vinhetas-âncora de personagens exibindo níveis baixos, médio e alto para a dimensão Conscienciosidade.....	15
Tabela 2- Exemplo de Vinhetas-âncora em polos opostos para o domínio Conscienciosidade.....	23
Tabela 3- Dicionário das Vinhetas-âncora com descrição dos itens e seu domínio, polo e nível	33
Tabela 4- Correlação das situações das Vinhetas-âncora com as notas escolares.....	45
Tabela 5- Correlação das situações das Vinhetas-âncora com as dimensões do SENNA	47
Tabela 6- Correlação das situações das Vinhetas-âncora com gênero e série escolar	49
Tabela 7- Correlação entre o índice global de consistência e os índices de violações das Vinhetas-âncora	50
Tabela 8- Correlação entre o índice de consistência global e os índices de violações com desempenho escolar.....	52
Tabela 9- Correlação entre os índices de aquiescência no traço para cada situação e geral.	54
Tabela 10- Correlação entre os índices de aquiescência no atributo para cada situação e geral.....	55
Tabela 11- Correlação entre os índices de aquiescência nos atributos e nos traços para cada situação e índices gerais de aquiescência.	56
Tabela 12- Correlações entre notas padronizadas de Linguagens e Matemática, gênero e séries escolares com os índices de aquiescência padronizados.	58
Tabela 13- Correlação de Pearson entre os índices de aquiescência no traço, atributo, global e as 60 dimensões socioemocionais	58
Tabela 14- Correlação entre os índices gerais e por dimensão de viés de grupo com notas em Linguagens e Matemática.....	64
Tabela 15- Correlação entre os índices gerais e por dimensão de viés de grupo com gênero, série escolar geral e séries escolares específicas.....	68
Tabela 16- Correlação entre os índices gerais e por dimensão de viés de grupo com as dimensões socioemocionais avaliados pelo SENNA.	69

Lista de abreviaturas e símbolos

ARS – Estilo de resposta aquiescente
ERS – Estilo de resposta extrema
MRS – Estilo de resposta central
SDR – Resposta socialmente desejável
GRB – Viés de grupo de referência
VA – Vinhetas-âncora

Apresentação

Esta tese teve como objetivo avaliar as evidências de validade para uso de Vinhetas âncora para mensuração e correção de erros do tipo sistemático causados por estilos de resposta em instrumentos de autorrelato, mais especificamente no instrumento SENNA v2.0. Este é um instrumento para avaliação das competências socioemocionais, dentro do modelo conhecido como *Big Five* Socioemocional, de crianças e adolescentes na faixa etária de 11 a 17 anos, e que estejam em ambiente escolar. Este instrumento foi desenvolvido por pesquisadores ligados ao Instituto Ayrton SENNA, no qual o orientador deste projeto faz parte desta equipe.

A necessidade de se investigar métodos e técnicas diferentes para mensurar e corrigir erros do tipo sistemático, em especial os causados pelos estilos de resposta, é uma constante na literatura psicométrica, com questionamentos acerca dos benefícios e malefícios sobre esta mensuração e correção. Erros do tipo sistemáticos são erros no quais os valores dos dados obtidos de uma amostra se desviam por um valor fixo dos valores verdadeiros. Desta forma, os estilos de resposta são considerados uma fonte do erro de tipo sistemático, sendo definidos como uma tendência em endossar os itens independente da condição do estímulo.

Este tema tem sido o foco do meu trabalho desde o mestrado, realizado no Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco, tendo trabalhado, de forma mais específica, com a resposta socialmente desejável (SDR). Durante esta etapa da minha pesquisa, pude perceber que por mais que os estudos caminhassem para desenvolver métodos e técnicas no controle dos estilos de resposta, muito pouco se tem sobre modelos que possam ser considerados como ‘padrão-ouro’ neste fim.

Com isto, espera-se contribuir não somente com a produção científica e acadêmica no que tange aos métodos psicométricos para mensuração e controle de estilos de respostas, mas de também possibilitar um refinamento teórico e experimental para a avaliação das

competências socioemocionais, que buscam realizar um diagnóstico das escolas públicas no país, contribuindo com a elaboração das políticas públicas da área.

Neste sentido, esta tese se encontra elaborado em algumas etapas. Na Introdução o leitor será apresentado de maneira breve sobre o problema de pesquisa. Em seguida, alguns tópicos foram trabalhados a fim de que se possa elaborar e justificar o problema de pesquisa, começando pelo problema de medida observada nos instrumentos de autorrelato. Com isto, uma revisão acerca dos estilos de respostas alvos de pesquisa deste projeto é apresentada e discutida, a fim de tornar o leitor mais familiarizado sobre o tema. Após esta revisão, a técnica das Vinhetas âncora é exposta, com exemplos de sua aplicação e formatação e de como este projeto buscará investigar as evidências de validade para a referida técnica. Finalizando a Introdução, uma breve explicação sobre o instrumento e suas bases teóricas são apresentadas. Com isto, os objetivos, hipóteses e os procedimentos metodológicos para a realização da pesquisa são descritos.

Em seguida, a sessão Resultados apresenta as análises realizadas e busca trazer o leitor para visualizar e acompanhar o passo a passo realizado desta pesquisa. Ainda, discute-se os resultados achados com base na literatura especializada, na sessão Discussão e se finaliza com um apanhado geral na parte de Considerações Finais.

Introdução

A avaliação por meio de instrumentos de autorrelato é amplamente utilizada em diversos contextos e para mensurar inúmeros construtos psicológicos. Dentre eles, ressalta-se aqui a avaliação das competências socioemocionais (SEMS), que baseia parte deste estudo. Na última década, alguns instrumentos foram propostos para a avaliação das SEMS, ou de construtos que foram pensados ser equivalentes, conforme demonstrou a revisão sistemática de Santos e Primi (2014). Contudo, o instrumento SENNA (Primi et al., 2021) tem apresentado maior robustez tanto no que tange o aspecto teórico das SEMS, quanto no aspecto metodológico para o controle de erros do tipo sistemático.

Mensurar e corrigir estilos de respostas, considerados como principal fonte de erros do tipo sistemático, é um problema ainda em debate na Psicometria, já que a distorção causada nos modelos testados poderia ser diminuída, ou até mesmo anulada, ao saber o grau do impacto dos estilos de resposta. Especialmente, instrumentos de autorrelato têm sido mais impactados por este tipo de erro, ainda que seja um dos métodos que melhor forneça informações acerca das diferenças individuais. Deste modo, pesquisas sobre as estimativas de precisão e as evidências de validade de instrumentos deste tipo buscam possibilidades de diminuir o impacto causado pelos estilos de respostas nestes parâmetros psicométricos supracitados.

Neste sentido, estudos prévios com uma técnica conhecida como Vinhetas-âncora (VA) puderam demonstrar benefícios do seu uso (cf. King et al., 2004; Primi et al., 2016a). Todavia, questionamentos mais recentes sobre o comportamento dos estilos de resposta em instrumentos de autorrelato, bem como modelos psicométricos mais refinados, colocam em xeque se esta técnica de fato pode ser benéfica (Möttus et al., 2012). Assim, o presente projeto se baseia na investigação das evidências de validade para o uso da técnica das VA em instrumentos de autorrelato, utilizando um instrumento para avaliação das competências socioemocionais como

ponto de partida (cf. Primi et al., 2016a; Van Vaerenbergh, & Thomas, 2013; von Davier et al., 2017; Weijters et al., 2010a).

Uso de instrumentos de autorrelato e o problema da sua medida

Nas ciências humanas e sociais, os pesquisadores geralmente usam instrumentos de autorrelato para mensurar traços latentes, tais como atitudes ou opiniões. Por ser de aplicação simples e até mesmo familiar aos entrevistados, bem como fáceis de avaliar e corrigir, este tipo de instrumento se tornou uma ferramenta popular de avaliação. Entretanto, sabe-se que os examinandos podem usar a escala de classificação deste tipo de instrumento de maneira diferente. Ou seja, eles interpretam as opções da escala de resposta de uma maneira própria e, com isso, diferem no uso de uma das opções de respostas em detrimento das demais, fazendo com que as respostas obtidas não reflitam de maneira fidedigna o traço latente mensurado (Berg & Collier, 1953; Cronbach, 1942; Danner et al., 2015; Primi et al., 2019).

Essa heterogeneidade no uso da categoria de resposta é chamada de estilos de resposta. Elas são amplamente definidas como preferências para categorias de respostas específicas que não estão diretamente relacionadas ao conteúdo do item (King et al., 2004; Nunnally, 1978; Paulhus, 1991; Vonkova et al., 2017). Isto acontece porque uma segunda dimensão, geralmente não correlacionada com o traço latente mensurado, está presente no momento do endosso ao item, fazendo com que os estilos de resposta possam ser considerados como uma fonte de erro do tipo sistemático (Cronbach, 1942; McCrae, 2018; Tracey, 2016).

Ao se responder um item, espera-se que a resposta obtida seja um reflexo do nível de traço latente do examinando. Contudo, pesquisas sobre os estilos de resposta destacaram que ao se endossar fortemente uma afirmação, este endosso pode ser o resultado de um nível moderado daquele construto em questão (como por exemplo, empatia) em combinação com uma tendência a respostas extremas. Os estilos de resposta mostraram-se quase sempre

presentes em avaliações por instrumentos de autorrelato, e o impacto causado por eles tem distorcido os modelos testados (Bolt & Johnson, 2009; Cronbach, 1942, 1946; Wetzel & Carstensen, 2017). Além disso, alguns estudos demonstram que os estilos de respostas tendem a ser estáveis e persistentes ao longo do tempo (Van Vaerenbergh & Thomas, 2013; Weijters et al., 2010a, 2010b; Wetzel et al., 2013; Wetzel et al., 2016).

Independentemente do conteúdo do item, alguns sujeitos preferem categorias extremas (Estilo de Resposta Extrema, ERS), enquanto outros podem preferir a categoria intermediária (Estilo de Resposta Intermediária, MRS), e, por fim, outros tendem a concordar independente do conteúdo do item, o que inclui ter o mesmo grau de concordância em itens que avaliam o mesmo traço latente, mas em polos opostos (Estilo de Resposta Aquiescente, ARS; ver Paulhus, 1991; Van Vaerenbergh & Thomas, 2013). Já a resposta socialmente desejável (SDR) ocorre quando um respondente compreende o conteúdo avaliado; entretanto, endossa o item da maneira como acredita que seja socialmente esperado (Cronbach, 1946; Falk & Cai, 2016).

O estilo de resposta conhecido como viés de grupo (ou *Group Reference Bias*, GRB) acontece quando uma pessoa se compara a um grupo de referência para dar sua resposta, de forma que poderá endossar mais um item ao se comparar com um grupo com habilidade menor que a sua, ou endossar menos se comparando com um grupo de maior habilidade (Marsh, 1987; Marsh & Parker, 1984). Desta forma, esta resposta dada não permite a discriminação do nível de traço latente do examinando por ele responder de acordo com uma referência que possui a um determinado grupo, confundindo a comparação das médias de grupos em um contexto avaliativo (Kyllonen & Bertling, 2014; Möttus et al., 2012).

Neste interim, quando os estilos de resposta não são levados em consideração, eles podem influenciar as propriedades psicométricas do instrumento, como a validade e precisão da medida. Uma resposta indicando grau de concordância alto com um determinado item pode ser o resultado de um alto nível naquele traço latente; mas também pode ser o resultado de um

nível moderado no traço latente em combinação com uma tendência a dar respostas extremas ou aquiescentes (e.g., Bolt et al., 2014; Plieninger, 2017; Van Vaerenbergh & Thomas, 2013).

Dos estilos de resposta já mencionados, ressalta-se que a literatura psicométrica tem destacado três estilos de resposta. São eles o estilo de resposta extrema (ERS), o estilo de resposta do ponto médio (MRS) e o estilo de resposta aquiescente (ARS). Em relação aos outros dois estilos de resposta, a saber resposta socialmente desejável (SDR) e viés de grupo (GRB), ainda que presentes na literatura, os pesquisadores os têm encarado de maneira distinta dos demais estilos mencionados. Em relação à SDR, não há consenso sobre sua causa e estabilidade nos sujeitos, de maneira que uma resposta socialmente desejável possa fazer parte da característica do sujeito (cf. McCrae, 2018; Tracey, 2016). As pesquisas sobre o GRB ainda são escassas, de modo que apenas em pesquisas transculturais este estilo de resposta tem sido abordado (Kyllonen & Bertling, 2014; Marsh, 1987).

Os estudos sobre esta fonte de erro do tipo sistemático não são apenas do interesse psicométrico, acadêmico, mas também social pelo impacto na elaboração de políticas públicas. Isto porque os estilos de resposta podem influenciar as comparações entre grupos. Ou seja, se uma subpopulação apresenta estilos de resposta diferentes de uma outra subpopulação (por exemplo, sexo feminino e masculino, ou alunos de escola pública ou particular), ao examinarmos as diferenças em um determinado traço de interesse tais diferenças não estarão somente ligadas aos traços em questão, mas também as diferenças de estilos de resposta entre subpopulações. Assim, as comparações entre estas subpopulações, ou grupos, podem ser tendenciosas, quando por exemplo, forem principalmente associadas aos estilos do que aos traços medidos ou quando uma diferença de estilo mascara uma diferença real no traço¹ (De Jong et al., 2008; Moors, 2004; van Herk et al., 2004).

¹ Esta situação pode também acontecer pelo DIF (funcionamento diferencial do item), que será abordado mais a frente nesta tese.

Para encontrar maneiras de lidar com estilos de resposta, os pesquisadores investigaram como o processo de mensuração pode ser modificado para reduzir estilos de resposta. Duas formas podem ser trazidas, a saber, por meios de métodos correlacionais ou experimentais (Lessa, 2018). Os métodos correlacionais se utilizam de técnicas estatísticas para chegar a um fator e, a partir dele, extrair a variância causada pelo estilo de resposta em questão – como a aquiescência ou resposta socialmente desejável. Já os métodos experimentais buscam realizar modificação na construção do instrumento, como uso de pares semânticos e alguns modelos no formato de escolha forçada (cf. Brown & Maydeu-Olivares, 2011). Apesar de todos estes esforços, os resultados observados, até aqui, têm sido inconsistentes (Böckenholt, 2017; Plieninger et al., 2019; Wetzel et al., 2016). Assim, a próxima sessão trará um breve descritivo sobre o desenvolvimento destes métodos de mensuração e controle dos estilos de respostas.

Desenvolvimento de métodos de mensuração e controle dos estilos de resposta

Inicialmente, Cronbach (1942)² e outros estudiosos tentaram medir os estilos de resposta usando estatísticas descritivas simples, como médias ou contagens entre itens. Por exemplo, Bachman e O'Malley (1984) calcularam o número de respostas extremas e usaram isso como uma medida do ERS, ou estilo de resposta extrema. Com o advento dos modelos da Teoria da Resposta ao Item (TRI), Análise Fatorial (AF) e Modelagem de Equações Estruturais (SEM), os pesquisadores observaram que uma possibilidade de lidar com os estilos de resposta seria encará-los como um fator adicional chamado fator de método, que consistiria em um segundo fator latente influenciando a resposta, mas irrelevante ao traço principal que se deseja medir (Falk & Cai, 2016; Primi et al., 2019; Wetzel et al., 2016).

² Ainda que os estudos iniciais possam ter sido realizados com técnicas estatísticas simples, o legado e os questionamentos deixados por Cronbach (1942, 1946, 1950) são ainda hoje utilizados para embasar pesquisas sobre o referido tema.

Os modelos de TRI foram desenvolvidos para representar a probabilidade de uma pessoa responder a um item, considerando os parâmetros do item e o nível de habilidade desta pessoa avaliada (DeMars, 2018; Hambleton & Swaminathan, 1985; Hambleton et al., 1991). Para a realização da técnica da TRI, é necessário que alguns pressupostos sejam observados, dentre eles a unidimensionalidade do item. Contudo, avanços metodológicos na TRI permitiram que mais de uma dimensão fosse testada no modelo, de forma a estimar como uma tendência de resposta a um determinado traço latente podem estar relacionados, em especial para os estilos de resposta de categorias extremas e centrais (ERS e MRS) (Böckenholt & Meiser, 2017; Meiser & Machunsky, 2008; Wetzal et al., 2013).

A técnica da análise fatorial estima a dimensionalidade do modelo por meio da covariância entre os itens. Uma contribuição desta técnica para os estilos de resposta se refere à aquiescência (ARS). Este estilo de resposta conseguiu ser encontrado em modelos de análise fatorial, pois ele tende a enviesar as correlações entre os itens quando a escala inclui itens positivos e negativos (Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006; Mirowsky e Ross, 1991; Zanon et al., 2018). Já na perspectiva das modelagens de equações estruturais (SEM) o estilo de resposta é encarado como uma variável latente para estimar sua variância no modelo, de forma a produzir um modelo o mais livre possível desta variável (Ferrando et al., 2016; Zanon et al., 2018). Vale ressaltar que estudos recentes têm buscado se utilizar de modelos mistos (*mixture models*) para compreender como o uso de uma escala, a partir de um viés de resposta, pode afetar essa estrutura psicométrica (Arias et al., 2020; Lubke and Muthén, 2007; Steinmann et al., 2021).

Os modelos psicométricos publicados geralmente incorporam vários estilos de resposta de uma só vez, ou seja, buscam investigar mais de um controle de estilo de resposta, como o ERS e MRS. Eles podem ser distinguidos em diferentes questões de pesquisa, (a) em relação ao modelo a ser utilizado enquanto a técnica estatística (por exemplo, modelo de créditos

parciais ou modelo de resposta nominal), e (b) se modelam estilos de resposta de forma exploratória ou confirmatória. Ainda, (c) se permitem ou não correlações no estilo de resposta referente ao conteúdo do item, e (d) se eles podem ser estimados pelas diferenças pessoais ou pelos parâmetros do item (Bolt & Johnson, 2009; Falk & Cai, 2016; Jin & Wang, 2014; Johnson & Bolt, 2010; Wang et al., 2006; Wetzel & Carstensen, 2017).

Plieninger (2017) coloca que a pesquisa sobre mensuração e controle nos estilos de resposta se concentra em três objetivos diferentes. Primeiro, em um nível psicométrico, em que modelos estatísticos são desenvolvidos para possibilitar a mensuração de estilos de resposta para se chegar a modelos que expliquem os dados de maneira parcimoniosa e teoricamente significativa. Segundo, em um nível mais substantivo, ou individual, interessando-se em descrever e explicar estilos de resposta, seja devido às diferenças individuais ou por uma questão cognitiva (neste último caso, a necessidade de se investigar o processo de resposta subjacente ao endosso de um item).

Ainda, chegando ao terceiro objetivo, considerado como nível aplicado, os pesquisadores não estão interessados em estilos de resposta *per se*, mas em retirar uma característica que leve a uma possível e potencial influência, considerada prejudicial, na medida do traço latente de interesse (Plieninger, 2017). Tais tentativas de se desenvolver um método capaz de mensurar e/ou controlar o erro do tipo sistemático pode ter feito com que a natureza da avaliação tenha sido perdida. Isto é, focou-se no método a partir de uma justificativa pelo próprio método, e não pela contribuição ao processo de avaliação (Falk & Cai, 2016; Tracey, 2016; Wetzel & Carstensen, 2017).

Estes três objetivos distintos acerca do mesmo tema, mensurar e controlar os estilos de resposta, levaram a uma quantidade considerável de estudos sobre o tema. Contudo, os resultados destes estudos pouco conseguem ser analisados para além dos seus objetivos, o que dificulta uma tomada de decisão acerca da viabilidade sobre seus impactos, seja na pesquisa ou

no campo prático (Plieninger, 2017; Henninger, 2019). Aqui, elenca-se três motivos que podem explicar o porquê de as pesquisas não conseguirem resultados mais robustos.

Primeiro, existe uma falta de resultados dos diferentes métodos em uma mesma amostra. Por exemplo, muitos métodos para controle da ARS foram propostos, e alguns destes estudos realizam a comparação de métodos para o controle da aquiescência (Falk & Cai, 2016; Maydeu-Olivares & Coffman, 2006; Zanon et al., 2018). Contudo, os resultados apresentados e discutidos são insipientes para eleger qual o melhor método possa ser utilizado para o controle da aquiescência, em especial por não considerar outros estilos de resposta, ou outras fontes de erros do tipo sistemático, nos dados utilizados (Ferrando et al., 2016; Johnson & Bolt, 2010; Kam & Zhou, 2015; Ten Berge, 1999; Wetzel & Carstensen, 2017).

O segundo motivo para a falta de resultados mais robustos está nos estudos que tenham investigado correlações no estilo de resposta com variáveis demográficas ou psicológicas, como traços de personalidade. Os resultados até aqui descritos pelos estudos não são consensuais quanto à eficácia do método empregado com o uso destas variáveis, deixando aberto se estes resultados foram de fato atingidos pelo método empregado, ou pelas diferenças normativas e de desenvolvimento. Assim, mesmo que os estilos de resposta possam ser medidos, as pesquisas ainda não foram capazes de propor um desenvolvimento de uma rede nomotética coerente em torno destas variáveis (Ferrando et al., 2016; Savalei & Falk, 2014; Soto et al., 2011; Wetzel & Carstensen, 2017).

Terceiro, a escolha de um método para medir e corrigir (ou, ao menos, controlar) os estilos de resposta é difícil para os pesquisadores porque faltam diretrizes concretas para subsidiar e até mesmo referenciar esta escolha. Por exemplo, na literatura psicométrica os modelos de variáveis latentes são muito populares, e esses modelos corrigem a característica do estilo de resposta como uma variável latente por meio de variáveis adicionais de estilo de resposta latente. Já para a psicologia transcultural, um método muito popular é a ipsatização,

também chamada de padronização intrassujeitos, que se baseia na correção das respostas observadas por meio da subtração da média de cada entrevistado (Fischer, 2004; Ten Berge, 1999; Soto et al., 2008). Mas a orientação sobre quando o método deve ser escolhido é escassa (cf. Savalei & Falk, 2014; Wetzel et al., 2016).

Ainda no caminho de estudos encontrados para controle de estilos de respostas, encontra-se dois estudos que merecem especial atenção por serem conhecidos como pioneiros nas técnicas experimentais e correlacionais da mensuração e controle de estilos de respostas. O primeiro, de caráter experimental, é o estudo de Peabody e Goldberg (1989), em que os autores propuseram que ter um conjunto balanceado de itens, isto é, itens de polo positivo e negativo (pares semânticos), trazem efeitos benéficos para a escala, haja vista que itens de polo positivo podem apresentar conteúdo mais valorativo, que seria propenso a ser endossado por um determinado estilo de resposta. Este método foi inicialmente pensado como forma de se controlar a resposta socialmente desejável, pois pares de itens semânticos conseguiriam atenuar o efeito de erro sistemático (McCrae, 2018; Peabody & Goldberg, 1989). Contudo, o uso deste tipo de pares também tem se mostrado satisfatório para o controle do estilo de resposta aquiescente e de tendência extrema (Batchelor & Miao, 2016; Kam, 2015; Primi et al., 2019).

O segundo estudo, correlacional, é o de Paulhus (1981), em que o autor propõe a criação de um componente a mais na estrutura do modelo para contabilizar a variância comum entre todos os itens do instrumento. Em outras palavras, é executada uma análise de componentes principais com todos os itens do questionário para que seja feita uma extração dos fatores “ $n + 1$ ”, onde n é o número esperado de fatores e 1 deve ser considerado como fator irrelevante para o modelo. Os resultados achados por Paulhus (1981) indicaram que ao estimar a variância deste fator, seguindo da sua exclusão, a comunalidade dos itens apresentaria um melhor ajuste. Contudo, este fator extra não permite identificar qual estilo de resposta está sendo mensurado. Assim, outros estudos buscaram refinar a ideia de contabilizar um fator geral para estimar um

estilo de resposta e, assim, controlar o impacto no modelo testado (cf. Soto et al., 2008; Savalei & Falk, 2014; Tracey, 2016; Wetzel et al., 2016).

Um ponto que permanece em aberto nas pesquisas sobre os estilos de respostas é se existem diferenças culturais que explicariam um funcionamento distinto para eles. Neste sentido, o estudo de He et al. (2015) analisou as respostas de 3.471 universitários de 20 países diferentes sobre o grau em que a resposta socialmente desejável (SDR) poderia impactar o endosso aos itens, de forma que uma análise de invariância, por meio da categoria estrutural. Os resultados demonstraram que a SDR está presente em nível semelhante independente do país em questão.

Um segundo ponto levanta a hipótese de que variáveis demográficas poderiam trazer uma tendência mais sistemática para endossar uma determinada opção na escala de resposta. Assim, fatores como sexo, idade e nível escolar poderiam explicar este uso de maneira mais consistente. Neste sentido, o estudo de Bachman et al. (2010) percebeu que pertencer a um grupo étnico, nos Estados Unidos, pode explicar um uso mais sistemático em relação a tendência extrema (ERS). No que tange ao estilo de resposta aquiescente (ARS), muitos estudos encontraram que uma capacidade cognitiva reduzida, em especial devido a um menor grau de escolaridade, ou menor habilidade cognitiva, constituem fatores determinantes para a ARS (Lechner & Rammstedt, 2015; McLarnon & Carswell, 2013; Vigil-Colet et al., 2012). Outros fatores, como sexo e idade, não encontram consenso na literatura sobre o quanto estão relacionados e impactam os estilos de respostas ARS, MRS e ERS (cf. Bachman & O'Malley, 1984; Savalei & Falk, 2016; Soto et al., 2008; Lelkes & Weiss, 2015).

Na contramão dos resultados supracitados, encontra-se os estudos de Bolt et al. (2014), Silva et al. (2018) e Stankov et al. (2017). Os resultados destes estudos demonstraram que, mesmo havendo o controle de características sociodemográficas, a forma de se utilizar uma escala de resposta pode apresentar uma variabilidade que não consegue ser explicada do ponto

de vista estatístico. Ou seja, a aplicação dos diferentes métodos já disponíveis para o controle dos estilos de respostas encontrados na literatura não apresentou ajustes mais satisfatórios. Tais resultados corroboram a justificativa de que as diferenças individuais têm um papel de extrema importância para o estudo dos estilos de respostas e que os métodos e técnicas até então desenvolvidos não conseguiram avançar neste tema.

Neste sentido, a revisão sistemática de Credé et al. (2010) examinou o impacto do efeito do grupo (GRB) em instrumentos de autorrelato. Os resultados concluíram que a pontuação de um indivíduo em um inventário é substancialmente influenciada pelo grupo de referência que ele possui, impactando significativamente nas evidências de validade baseadas em critérios externos, neste caso, as médias escolares produzidas pelo rendimento acadêmico ao longo do ensino superior. Algo semelhante tem sido debatido nos resultados do PISA (OECD, 2014), uma vez que os resultados obtidos nos construtos avaliados por instrumentos de autorrelato (e.g., interesse em Matemática) apresentaram discrepância entre o rendimento na prova (nível individual) e os resultados apresentados pelo país (*aggregated country level*).

Em resumo, pesquisas anteriores fizeram progressos em termos de mensuração de estilos de resposta, e modelos abrangentes e flexíveis estão disponíveis hoje. Além disso, o campo de investigação sobre estilos de respostas concretizou a existência e os impactos dos estilos de resposta, em especial dos ERS, ARS e MRS, e que pode haver benefícios em controlá-los. Isto porque conseguir estimar o traço latente de um indivíduo, ou grupo, de maneira mais precisa e fidedigna permite que intervenções e até mesmo políticas públicas possam ser elaboradas com maior poder de decisão, beneficiando à sociedade de maneira geral. Além disso, existe um grande interesse em correlatos de estilos de resposta, como o funcionamento diferencial do item (DIF), especialmente em estudos interculturais (OECD, 2014; McCrae, 2018; Primi et al., 2019; Tracey, 2016; Wetzel et al., 2016).

No entanto, permanecem os caminhos para futuras pesquisas e algumas questões específicas são abordadas no presente projeto. Como dito, três estilos de respostas, a saber tendência de resposta extrema (ERS), estilo de resposta aquiescente (ARS) e tendência de resposta do meio (MRS), ganham destaque na literatura e métodos são desenvolvidos com robustez para mensuração e controle. Já os estilos de respostas como a SDR e o viés de grupo (GRB) ainda têm um longo caminho de estudos para ser feito e angariar evidências de validade de que sua mensuração e controle também trazem benefícios (ou não) ao campo aplicado. Assim, a técnica das Vinhetas-âncora foi escolhida, como objeto desta tesa, enquanto uma possibilidade de mensurar e corrigir estilos de respostas.

Técnica das Vinhetas-âncora para mensuração e correção de estilos de respostas

A literatura fora da área da Psicologia, como Ciências Políticas e Saúde Pública, tem compreendido os estilos de resposta, e conseqüentemente as fontes de erro sistemático, como o funcionamento diferencial do item, conhecido como DIF (cf. King et al., 2004; Silva et al., 2018). Isto porque o DIF ocorre quando pessoas de diferentes grupos, mas com o mesmo grau de traço latente, dão respostas diferentes, não podendo ser possível afirmar que o item, ou itens, é invariante (Karami, 2012; Thissen et al., 1993). Contudo, Johanson e Osborn (2004) propõem que o estilo de resposta aquiescente pode ser compreendido como funcionamento diferencial da pessoa, ou DPF.

De forma distinta ao DIF, o DPF acontece quando dois itens de polos distintos (i.e., negativo e positivo) têm igual nível de dificuldade e se propõem a medir um mesmo construto, apresentam diferentes probabilidades de endosso, isto é, $P(X_{ij_1}|\theta) \neq P(X_{ij_2}|\theta)$, de forma que P é a probabilidade de resposta X ao item j , pelo examinando i ao item positivo 1 ou negativo 2 (Johnason & Osborn, 2004; Primi et al., 2019). Um exemplo de DIF seria o exposto: uma prova de história do Brasil é elaborada para alunos do 8º ano do Ensino Fundamental II de uma

determinada escola. Caso haja alunos estrangeiros, fazendo intercâmbio nesta escola, ainda que tenham o mesmo conhecimento de história que os alunos brasileiros, os itens desta prova de história poderão ser mais difíceis para os alunos estrangeiros do que para os alunos brasileiros em razão das diferenças no conhecimento da língua e não do conhecimento de história.

Em relação ao DPF, pode-se exemplificar da seguinte maneira: em um questionário de traços de personalidade, ao estar diante de um item como “sou amável e legal com quase todo mundo” (polo positivo; Primi et al., 2016a), uma pessoa endossa a categoria de resposta quatro, numa escala do tipo Likert de 5 pontos. Para o item oposto, “é difícil para mim dar atenção às pessoas” (polo negativo, Primi et al., 2016), pressupondo que meça o mesmo construto e tenha o mesmo nível de dificuldade do primeiro o esperado é que a resposta dada seja a categoria dois. Se a pessoa, diferente do esperado, endossa a categoria três ou superior estamos diante de DPF, já que a pessoa com um nível de *theta* fixo tem uma resposta diferente a itens similares. Assim, uma forma de se explicar os estilos de respostas é por meio da concepção do DPF, já que o estilo de resposta é uma característica da pessoa e não do item o que justifica, a partir daqui utilizar-se o termo DPF, e não DIF (Johnason & Osborn, 2004; Primi et al., 2019).

Uma técnica desenvolvida no âmbito das ciências sociais e políticas por King et al. (2004) para controlar o DPF é a Vinheta-âncora (VA). Ela realiza descrições curtas de situações fictícias, de forma que o indivíduo é requerido para classificá-las usando questões típicas de autorrelato com a mesma escala de resposta, mas nesse momento, ele tem o papel de responder sobre o personagem da vinheta. (Hopkins & King, 2010). Em outras palavras, as VA são textos breves que descrevem um (ou mais) personagem em terceira pessoa exibindo um nível determinado (geralmente baixo, médio ou alto) no traço latente na característica avaliada no instrumento de autorrelato (Hopkins & King, 2010). Com isto, o uso desta técnica permite medir como o examinando usa a escala de resposta, por exemplo, como usos extremos ou se

usa diferentes *benchmarks*³ de comparação para julgar o nível dos traços. Sendo possível medir os estilos de resposta, seria possível também identificar efeitos que algumas covariáveis (e.g., nível socioeconômico, escolaridade etc.) nesses estilos (He & van de Vijer, 2016, Yousfi, 2018).

Para ilustrar, pode-se observar o seguinte caso: “Aline costuma deixar as coisas todas bagunçadas, odeia limpar a casa e deixa as lições de casa sem completar. Quanto você acha que Aline é organizada?” Esta situação descreve uma pessoa com um baixo nível de comportamento organizado, sendo esperado que este item seja endossado nas opções 1 ou 2 (numa escala tipo Likert de 5 pontos). Se, para esta situação, uma pessoa responde qualquer valor diferente de 1 ou 2, ela está respondendo de maneira inconsistente com o esperado (King et al., 2004). Assim, a partir desta incongruência, é possível aprender como o sujeito usa a escala de resposta, possibilitando um eventual ajuste das suas respostas aos itens e chegar a escores mais fidedignos quanto àquele comportamento avaliado. A Tabela 1 traz alguns exemplos de Vinhetas-âncora.

Tabela 1

Exemplo de três Vinhetas-âncora de personagens exibindo níveis baixos, médio e alto para a dimensão Conscienciosidade

Aline costuma deixar as coisas todas bagunçadas, odeia limpar a casa e deixa as lições de casa sem completar. Quanto você acha que Aline é organizada?
Nada organizada 1 ----- 2 -----3 -----4 -----5 Completamente organizada
Manuela tem noção de organização, mas às vezes deixa seu quarto bagunçado. Tende a completar as lições em cima da hora. Quanto você acha que Manuela é organizada?
Nada organizada 1 ----- 2 -----3 -----4 -----5 Completamente organizada
Juliana é bastante cuidadosa e dedicada. Gosta de limpar a casa, é caprichosa nas lições e sempre as termina antes do prazo. Quanto você acha que Juliana é organizada?
Nada organizada 1 ----- 2 -----3 -----4 -----5 Completamente organizada

Fonte: Adaptado de Primi et al. (2016a)

³ Pode-se traduzir este termo para ‘pontos de referência’, no contexto em que se está utilizando-o.

O uso das Vinhetas-âncora pode trazer vantagens para as pesquisas quantitativas. A primeira vantagem elucidada é o fato de que seu formato é em itens, fazendo com que os respondentes não ativem processos cognitivos de bloqueio de respostas, ou de falseamento (He & van de Vijer, 2017). Pelo contrário, adicionar os itens das vinhetas nas avaliações pode fazer com que os respondentes possam se engajar mais na tarefa e trazer respostas mais legítimas ao estudo em questão. Um outro ponto é o baixo custo que o uso de Vinhetas-âncora apresenta. No mais, as VA podem ser utilizadas tanto em pesquisas com número de participantes reduzidos (i.e., $n = 50$), ou em larga escala (e.g. *World Health Survey*, *Programme for International Student Assessment* [PISA] etc.), possibilitando resultados mais concretos para a elaboração de políticas públicas (He & van de Vijer, 2016). Por outro lado, um dos pressupostos básicos para uso das Vinhetas-âncora é chamado de *Consistência*, isso é, as vinhetas só funcionam se o indivíduo utilizar a escala de respostas às VA da mesma forma como utiliza para responder aos itens do questionário (Hopkins & King, 2010; Tandon et al., 2003). Em outras palavras, ele precisa responder as vinhetas da mesma maneira que ele responde aos questionários. Violações desses pressupostos colocam em xeque a validade de construto da vinhetas para medir estilos de resposta. Assim tornava-se necessário fazer esta investigação quanto à violação sempre que um conjunto das VA era utilizada em paralelo a um questionário (He & van de Vijer, 2017; Tandon et al., 2003).

O uso prático das vinhetas é feito por meio da recodificação não-paramétrica das respostas ao questionário. Wand et al. (2011) desenvolveram um pacote *anchors* para o ambiente R de programação (*R Development Team*, 2021) a fim de realizar essas recodificações e diagnósticos das respostas às vinhetas. Em um primeiro momento as respostas dadas nas vinhetas são corrigidas para verificar se as respostas estão congruentes ou incongruentes, ou seja, se as respostas dadas para as situações hipotéticas estão dentro do que se espera para cada um dos casos apresentados. Esta situação é chamada de verificação se o pressuposto de

ancoragem das vinhetas foi alcançado (Kyllonen & Bertling, 2014; Wand et al., 2011). Em outras palavras, é esperado que (a) haja três respostas diferentes para cada situação hipotética dado pelo sujeito (y_s) e (b) que as respostas respeitem a ordenação de escores baixo, médio e alto, de forma que a resposta um (y_{s_1}) tenha um valor menor que a resposta dois (y_{s_2}), que por sua vez seja menor que a resposta três (y_{s_3}), apresentando o seguinte ordenamento: $y_{s_1} < y_{s_2} < y_{s_3}$.

Kyllonen e Bertling (2014) desenvolveram um método específico para correção de Vinhetas-âncora para uso de instrumentos em escala do tipo Likert, tomando como base os estudos sobre a correção não paramétrica de King et al. (2004) e Wand et al. (2011). Este método recodifica as respostas dadas no instrumento de autorrelato para uma nova pontuação, a partir das respostas dadas nas Vinhetas-âncora. Para definir a quantidade de pontos que esta escala terá, utiliza-se a fórmula $2k + 1 = X$, de forma que k é a quantidade de Vinhetas-âncora usadas por conjunto/domínio e X será a nova escala de pontuação. Desta forma, ao pensar em um conjunto de VA que possui três itens, a nova pontuação desta escala será de 7 pontos (Kyllonen & Bertling, 2014). A Figura 1 ilustra o processo de correção pelas Vinhetas-âncora.

Figura 1

Regras para correção não paramétrica das vinhetas-âncora em um conjunto de três situações hipotéticas

Regras de recodificação	$y_s < v_{lo}$	$y_s = v_{lo}$	$v_{lo} < y_s < v_{md}$	$y_s = v_{md}$	$v_{md} < y_s < v_{hi}$	$y_s = v_{hi}$	$y_s > v_{hi}$
Exemplo 1	<p>Vinhetas</p> <p>Recodificação</p>						
Exemplo 2	<p>Vinhetas</p> <p>Recodificação</p>						
Regras de recodificação	$y_s < v_{lo}$	$y_s = v_{lo}$	$v_{lo} < y_s < v_{md}$	$y_s = v_{md}$	$v_{md} < y_s < v_{hi}$	$y_s = v_{hi}$	$y_s > v_{hi}$
Exemplo 1	v_{lo}	v_{md}	v_{hi}				
Exemplo 2	v_{lo}	v_{md}	v_{hi}				
Exemplo 1	y_s^*						
Exemplo 2	y_s^*						
Exemplo 1	1	2	3	4	5	6	7
Exemplo 2	1	2	3	4	5	6	7

Nota. Exemplo 1 para apenas uma resposta a ser recodificada. Exemplo 2 para 3 respostas a serem recodificadas.
Fonte: Adaptado de Kyllonen e Berg (2014) e Primi et al. (2016a).

No que se mostra na Figura 1, a escala original possui uma gradação de 5 pontos, em que 1 pode ser ‘não tem nada a ver com a pessoa’ e 5 ‘tem tudo a ver com a pessoa’. O exemplo 1 traz a situação em que a pessoa respondeu 1, ‘não tem nada a ver com a pessoa’, para a Vinheta-âncora de traço baixo (v_{lo}). Como a resposta dada foi igual a resposta esperada, ou seja, $y_s = v_{lo}$, o escore da pessoa passa a ser 2. No exemplo 2, a resposta dada para a VA de traço médio (v_{md}) foi menor que o esperado, resultando em um novo escore de 3. Por fim, ainda no exemplo 2, a resposta para a VA de traço alto (v_{hi}) foi 4, sendo menor que o esperado, gerando um novo escore de 5.

Muitos estudos de áreas como Saúde Pública e Ciência Política têm mostrado resultados satisfatórios com a utilização da técnica das Vinhetas-âncora. Por exemplo, os estudos de Crane et al. (2015) e Léon et al. (2014) demonstraram que o uso desta técnica conseguiu mensurar e corrigir as diferentes formas de uso da escala de resposta a partir de grupos com caracterização demográfica distintas. Ainda, os resultados de Hirve et al. (2014) e Weiss & Roberts (2018) concluíram que os modelos testados tiveram um melhor ajuste, além de oferecer uma solução acessível para os problemas de comparação entre grupos que apresentem um funcionamento diferencial ao item (DIF), ou funcionamento diferencial da pessoa (DPF).

Diante disto, este ajuste poderia ser uma possibilidade de se corrigir respostas dadas em um instrumento de autorrelato que estão sendo impactados por um estilo de resposta, como explicação para o DPF (King et al., 2004; Kyllonen & Bertling, 2014; Primi et al., 2019). Uma revisão de literatura sobre o tema foi realizada ainda no delineamento do presente projeto, em fevereiro de 2019, e utilizou-se das bases de dados *DataSearch*, *PsycINFO*, *PubMed Central*, *Redalyc* e o Portal de Periódicos da CAPES (que é uma base de dados que visa disponibilizar acesso a pesquisadores brasileiros para diversas bases de dados internacionais). Não houve limitação de ano, idioma ou área de estudo.

Nas bases *DataSearch* e *PubMed Central* as palavras usadas para a recuperação dos estudos foram “*anchoring AND vignettes*” AND “*common AND method**”. Nas bases *Redalyc* e *PsycINFO* optou-se apenas a utilização das palavras-chaves “*anchoring AND vignettes*”, uma vez que ambas as bases o uso dos demais descritores apresentaram poucos estudos. Já no Portal de Periódico CAPES, utilizou-se as mesmas palavras-chaves das bases *DataSearch* e *PubMed Central*. Contudo, outros filtros, do próprio portal, foram utilizados, a saber, “artigos”, “revisado por pares”, “*education*”, “*psychology*”, “*vignettes*”, “*Anchoring vignettes*”, “*self-rating*”, “DIF”, “*Reporting heterogeneity*”.

Com os resultados das buscas, os estudos tiveram seus títulos e resumos lidos para uma prévia seleção. Artigos não empíricos, como revisão de literatura e artigos teóricos, não foram considerados. Após esta etapa, os artigos foram lidos na íntegra. Nesta fase, excluiu-se os artigos que não tratavam sobre o uso de Vinhetas-âncora e estudos que, no decorrer de suas análises, optaram por não utilizar as VA em seus resultados. Com os artigos selecionados, uma lista foi feita no *software* de planilha de dados *Microsoft Excel®*, de forma que cada artigo ocupe uma linha, e cada coluna indique informações específicas de cada um deles. As informações foram: autor, ano, tipo de pesquisa, objetivo, tipo de amostra, instrumentos utilizados, resultados e limitações de cada estudo. Estes dados permitiram uma análise sistemática acerca do uso das VA em cada um dos estudos,

Os resultados evidenciam que diferentes áreas temáticas se utilizam desta técnica e que seu uso tem demonstrado efeitos satisfatórios no uso do método de autorrelato. Estes achados estão consonantes com o estudo de Hopkins e King (2010), ao enfatizar que as VA poderiam ter usufruto em diferentes aplicações. Isto porque o método de autorrelato é utilizado em diferentes delineamentos de pesquisa, em especial por diversos países para monitoramento de variáveis para construção de políticas públicas, o que torna o uso de Vinhetas-âncora necessárias para que as diferenças observadas entre e dentre grupos possam ser controladas, resultando em escores mais fidedignos (Angelini et al., 2012; He & van de Vijer, 2016, Primi et al., 2016).

Ademais, atesta-se com estes resultados àqueles encontrados por He et al. (2016) e Weiss e Robert (2018). Neles, é possível observar que os índices de ajustes e outras evidências de validade também são mais bem relacionadas a partir do ajustamento do escore. Ainda, o estudo de Kyllonen e Bertling (2014), com os dados do PISA 2012, levou a uma observação de que a correção dos instrumentos de autorrelato pelas Vinhetas-âncora possibilitou uma melhor compreensão em como alunos de diferentes países e culturas respondem aos mesmos itens de

maneira diferente. Vale ressaltar que o escore de um respondente é, no método de autorrelato, seu desempenho – e é a partir deste desempenho que se pode realizar inferências em relação ao traço latente/comportamento que se pretende medir (Cronbach, 1946; Messick, 1980).

Na contramão de evidências favoráveis para o uso da VA, estão os estudos de Möttus et al. (2012) e Von Davier et al. (2017). Ao corrigir as respostas de estudantes universitários de mais de 21 países no domínio Conscienciosidade, os resultados apresentados por Möttus et al. (2012) mostraram que o uso da correção por VA não apresentou diferença estatisticamente significativa nas médias não corrigidas e corrigidas. Já os resultados de von Davier et al. (2017) questionou se a melhora nos resultados apresentados por Kyllonen e Bertling (2014) não se deve ao fato de que o processo de recodificação das respostas pode ter sido devido a uma melhora nos índices de precisão. Isto porque as respostas ficaram mais homogêneas – inflando a correlação entre os itens e impactando estes índices, mas não trazendo benefícios para evidências de validade baseadas em critério externo (von Davier et al., 2017; Primi et al. 2016).

Dada a falta de consenso no uso da técnica de Vinhetas-âncora para um impacto positivo em instrumentos de autorrelato no campo da Psicometria para correção de erros do tipo sistemático, o presente estudo é proposto a afim de realizar uma análise inicial de validade desta técnica. Mais especificamente, realizou-se várias análises exploratórias e de correlação com variáveis externas procurando responder uma pergunta fundamental: o que as Vinhetas-âncora medem? Esta pergunta é feita ao observar que os vários estudos revisados focam no uso das vinhetas em aplicações práticas recodificando as repostas com a esperança de que consigam purificar os escores de autorrelato. Mas poucos estudos buscam avaliar os pressupostos subjacentes as vinhetas. Considera-se que saber se o quê as VA conseguem de fato medir é o primeiro passo para compreender seu uso dentro deste campo.

A segunda pergunta está na indagação de que se as VA eventualmente conseguirem medir estilos de resposta, quais estilos de resposta elas conseguem medir? Sabendo-se que a

heterogeneidade no uso das escalas de resposta pode se dar devido a resposta aquiescente (ARS), estilo de resposta dos pontos extremos (ERS), estilo de resposta do ponto central (MRS) e grupo de referência (GRB), qual desses estilos as Vinhetas-âncora conseguiriam medir? É preciso ressaltar que condições diversas, como capacidade cognitiva, variáveis demográficas, a saber sexo, escolaridade, etnia etc., foram devidamente testadas com a técnica de Vinhetas-âncora, sendo assim, elas conseguiriam medir estilos de resposta como o estilo de resposta aquiescente (ARS), estilo de resposta dos pontos extremos (ERS), estilo de resposta do ponto central (MRS) e grupo de referência (GRB)?

Destarte, nenhum dos estudos revisados incluíam itens de polos negativos (representando níveis baixos do traço) como parte das Vinhetas-âncora. Estudar conjuntos de Vinhetas-âncora no sentido do polo negativo poderia contribuir para o estudo do uso desta técnica na mensuração de erros sistemáticos, e de maneira mais específica, estilos de resposta. Isto porque um conjunto de VA, no polo positivo, fornece informações sobre o endosso em cada um dos itens nos determinados níveis de habilidade da situação hipotética. Assim, um conjunto de VA do mesmo domínio no polo negativo seria esperado que tivesse o endosso inverso, sendo qualquer diferença entre estas pontuações uma possível evidência de erro sistemático (Ten Berge, 1999; Primi et al., 2019). Assim esta tese contribui estudando pela primeira vez Vinhetas-âncora com itens de polos negativos em adição dos itens com polos positivos.

Para responder tais perguntas e elaborar formas de superar as limitações até aqui encontradas nos resultados sobre o uso da técnica de Vinhetas-âncora, alguns pontos foram analisados. O primeiro ponto diz respeito à construção das vinhetas em si. Sabe-se que usar um mesmo conjunto de vinhetas para diversos construtos não é o ideal (cf. Grol-Prokopczyk; 2018; Hopkins & King, 2010; Knott et al., 2017). O segundo ponto é referente ao conhecimento já produzido para o controle de estilos de respostas, a saber, o uso de pares semânticos (Arnulf et

al., 2018; Peabody & Goldberg, 1989; Primi et al., 2019). Neste sentido, vinhetas construídas em polos opostos possibilitariam acessar mais informações sobre a maneira como um indivíduo está usando a escala de resposta. A Tabela 2 exemplifica uma situação de VA construída em par semântico, isto é, em polos opostos.

Tabela 2

Exemplo de Vinhetas-âncora em polos opostos para o domínio Conscienciosidade

Aline costuma deixar as coisas todas bagunçadas, odeia limpar a casa e deixa as lições de casa sem completar.
Quanto você acha que Aline é organizada?

Nada organizada 1 ---- 2 ----3 ----4 ----5 Completamente organizada

Aline costuma deixar as coisas todas bagunçadas, odeia limpar a casa e deixa as lições de casa sem completar.
Quanto você acha que Aline é bagunceira e desorganizada?

Nada desorganizada 1 ---- 2 ----3 ----4 ----5 Completamente desorganizada

Fonte: Adaptado de Primi et al. (2016a)

A hipótese central do presente estudo é que a técnica das Vinhetas-âncora é útil para mensurar estilos de resposta em avaliações baseadas no método do autorrelato. Ao analisar a maneira como esta técnica funciona, pode-se inferir que qualquer variação nas respostas às vinhetas representa diferentes interpretações dos níveis do traço que deveriam ter poucas variações. Assim, hipotetiza-se que variações nas respostas às vinhetas refletiriam estilos de resposta (King et al., 2004; Primi et al., 2019).

Esta tese foca em um inventário para avaliação das competências socioemocionais para adolescentes em ambiente de ensino formal, o SENNA (Primi et al., 2016a; Santos & Primi, 2014). Este instrumento possui um conjunto de Vinhetas-âncora no sentido do polo positivo para as cinco dimensões do modelo *Big Five*, bem como um conjunto destas mesmas VA no sentido do polo negativo destas dimensões (Primi et al., 2016a, 2019). O tema das competências socioemocionais tem ganhado destaque na literatura internacional, e, assim, espera-se que possibilitar mais evidências de validade e avaliar as estimativas de fidedignidade deste

instrumento com o uso das VA trará benefícios não somente científicos, mas também sociais, como explanado a seguir.

Avaliação das competências socioemocionais por instrumento de autorrelato

Nas últimas duas décadas, as ciências da educação e as políticas de educação pública desenvolveram um interesse na avaliação e aprendizagem de habilidades socioemocionais, também conhecidas como habilidades do século 21 (a partir daqui, abreviado como SEMS). Este interesse se baseia na noção de que somente o aprendizado de habilidades cognitivas para o ambiente escolar (como Matemática e Linguagens) não é o suficiente para uma vida de sucesso pessoal e profissional (John & Fruyt, 2015).

Atualmente, além de indicadores de desempenho escolar mais tradicionais (como notas de linguagem e Matemática), as SEMS também são vistas como um produto do ambiente e do processo educacional. Com mais atenção às SEMS, traz-se questionamentos sobre seu *status* conceitual, sua estrutura básica e como elas se desenvolvem – se naturalmente, ou seja, seu desenvolvimento normativo, ou se é socialmente influenciado por experiências formais e informais da intervenção escolar (Durlak et al., 2011; John & Fruit, 2015; Kyllonen et al., 2014; Roberts et al., 2006).

Buscar uma definição sobre as SEMS tem perpassado muitos grupos, bem como formas de medidas, intervenções e aplicações em políticas educacionais. Todos estes grupos, ainda que divergentes em algum aspecto, partem do conceito em que adquirir e aplicar uma determinada competência é essencial para (a) reconhecer e gerenciar emoções; (b) estabelecer e alcançar objetivos positivos; (c) apreciar as perspectivas dos outros; (d) estabelecer e manter relacionamentos de apoio; (e) tomar decisões responsáveis; e (f) lidar com situações pessoais e interpessoais de forma construtiva (Duckworth & Yeager, 2015; Osher et al., 2016).

Diante do arcabouço teórico encontrado para definir as SEMS, o presente projeto focará no que é proposto por De Fruyt et al. (2015). Segundo os autores, essas habilidades são padrões consistentes de pensamentos, sentimentos e comportamentos, tornando-se importantes para os desfechos econômicos e sociais ao longo da vida de um indivíduo. Contudo, estes padrões são maleáveis e são, potencialmente, impactados por meio de experiências de aprendizagem formal e informal. Esta definição se aproxima bastante nos modelos fatoriais de personalidade, mais especificamente com o modelo dos cinco grandes fatores (*Big Five*), fazendo com que as SEMS sejam também conhecidas como *Big Five* Socioemocional (cf. Abrahams et al., 2019; De Fruyt et al., 2015; John & De Fruyt, 2015; Primi et al., 2021).

Concomitante a estas investigações científicas, intensos debates sobre políticas educacionais em muitos países se iniciaram, incluindo o Brasil, sobre como representar a aprendizagem das SEMS no currículo educacional (OECD, 2015). Uma iniciativa significativa no Brasil é a Base Nacional Curricular Comum (BNCC; MEC, 2017). Nas discussões para o desenvolvimento da BNCC, tornou-se evidente a necessidade de uma taxonomia na qual possa se basear e que articule com os achados científicos das SEMS, já que demonstraram ser importantes para a aprendizagem e o funcionamento no século XXI (MEC, 2017; OCDE, 2015).

O interesse tanto na avaliação quanto no desenvolvimento das SEMS vem crescendo na última década, como uma forma de melhorar o bem-estar geral e preparar os jovens para os desafios futuros por meio de programas de educação e intervenção (OECD, 2015). Primi e Santos (2014) iniciaram uma busca sistemática e identificaram oito instrumentos que avaliavam os construtos das SEMS com o poder de prever resultados consequentes da educação, sendo estes instrumentos viáveis de administrar (isto é, possuíam baixo custo para sua aplicação, sua administração demandava pouco tempo, os itens eram formulados em linguagem apropriada para a idade, além de possuir instruções claras), avaliando construtos maleáveis e demonstrando características psicométricas robustas. Os resultados do primeiro estudo de Santos e Primi

(2014), quanto os resultados de Primi et al. (2016b), demonstraram que os itens e escalas dessas oito medidas frequentemente usadas para avaliar a aprendizagem das SEMS poderiam ser facilmente mapeados no esquema do *Big Five*.

Assim, a proposta do instrumento para avaliar as SEMS encontra-se na sua segunda versão (SENNA v2.0; Primi et al., no prelo). Esta versão do SENNA abarca as cinco dimensões do *Big Five*, a saber, Extroversão (chamado no SENNA v2.0 de Engajamento com os outros), Neuroticismo (Resiliência Emocional, no SENNA v2.0), Abertura (Abertura à novas ideias, SENNA v2.0), Conscienciosidade (Autogestão, SENNA v2.0) e Agradabilidade (Amabilidade, SENNA v2.0). Estas dimensões são caracterizadas por conseguir ter uma explicação maior acerca dos traços de personalidade, além de ter a replicação do modelo em diversos estudos de diferentes países e em situações distintas (cf. Digman, 2002; John et al., 2008; Poropat, 2009; OECD, 2015; Soto & John, 2017).

Uma segunda vantagem deste instrumento é que a sua estrutura fatorial, ou principais dimensões, são compostos por facetas, que são um nível mais específico na explicação das características dos traços de personalidade (ver Soto & John, 2017). No SENNA v2.0, as facetas conseguem fomentar uma descrição mais rica e compreensiva das características socioemocionais daqueles que respondem ao instrumento (Primi et al., 2019). Ainda sobre o instrumento, é necessário enfatizar que sua construção engloba itens elaborados como pares semânticos, isto é, itens elaborados para mensurar a faceta Compaixão (pertencente à dimensão Amabilidade) possui três itens de polo positivo, e três itens de polo negativo. Ainda, cada faceta possui três itens de autoeficácia, que buscam complementar a avaliação das características socioemocionais a partir das crenças dos adolescentes sobre suas próprias habilidades (Primi et al., 2019)

Estudos de evidências de validade do SENNA v2.0 estão em andamento e os resultados de alguns destes estudos já se encontram disponíveis e apresentaram-se como satisfatórios (cf.

Abrahams et al., 2019; Primi et al., 2015, 2016b, 2021). Contudo, apenas o estudo de Primi et al. (2016a) chegou a avaliar o impacto do uso de Vinhetas-âncora no presente instrumento. Nele, os resultados demonstraram que a correção das respostas pelas VA, como controle da aquiescência, melhorou os índices de precisão do instrumento, mas que as evidências de validade critério (isto é, as notas escolares em Linguagem e Matemática) tiveram uma diminuição no valor preditivo. Este resultado pode ser um indício de que o processo de resposta, tanto para as situações hipotéticas das VA, quanto para a autoavaliação nos itens precisa ser mais bem investigado para o controle de estilos de respostas.

Diante da escassez de pesquisas sobre o tema, tanto sobre a necessidade de mensurar e controlar estilos de respostas em instrumentos de autorrelato no país, bem como as propriedades psicométricas do SENNA v2.0 com o uso de Vinhetas-âncora, apresenta-se esta tese. Conforme exposto, esta tese, então, buscará explorar os seguintes pontos, a saber (a) averiguar que estilos de respostas, dentre ERS, ARS, MRS e GRB, podem ser mensurados e controlados com o uso da técnica das Vinhetas-âncora, (b) de que forma seu uso está relacionado com o processo de resposta no método do autorrelato a partir das relações com as variáveis externas, em especial as de cunho cognitivo (desempenho) e características sociodemográficas. Por fim, com base nestes resultados, (c) propõe-se criar uma ferramenta que auxilie na correção desta técnica em pesquisas e na prática profissional, mostrando aos pesquisadores e profissionais índices de estilos de resposta presente nos seus dados. É importante ressaltar que esta ferramenta é um fruto de inovação científica e metodológica, pois se baseia nos resultados dos índices de mensuração e controle para estilos de resposta, com base na sensibilidade que os processos cognitivos demonstram no endosso ao estilo de resposta.

Objetivo

O objetivo geral desta tese é investigar a validade das Vinhetas-âncora como medida de estilos de resposta. A fim de que este objetivo seja alcançado, os seguintes objetivos específicos foram elaborados, sendo dividida em quatro partes. A primeira parte do estudo visa realizar uma análise descritiva exploratória das respostas às Vinhetas-âncora, buscando compreender a forma como as escalas de respostas das VAs são utilizadas. Na segunda parte, avaliar-se-á as evidências de validade baseadas na relação com variáveis externas nomeadamente com características de personalidade (dimensões do *Big five* socioemocional) e capacidade cognitiva (desempenho acadêmico em português e Matemática). A terceira parte investiga se é possível mensurar aquiescência a partir das Vinhetas-âncora, verificando-se a correlação desse índice com variáveis externas e com medidas de aquiescência do SENNA. De forma semelhante, a quarta parte investigará a validade para as VAs mensurarem o estilo de resposta de viés de grupo.

Hipóteses

As hipóteses que guiam esta tese estão divididas de acordo com os objetivos propostos e em cada estudo. Vale salientar que o norte desta tese está em avançar na compreensão de como se dá o funcionamento individual para o endosso a um, ou mais, estilo de resposta. A primeira hipótese é de que a técnica das VAs pode nos ajudar a compreender como o uso de uma escala de resposta pode estar associado com o tipo de item (i.e., item de polo positivo ou de polo negativo) que se aplica em um instrumento de autorrelato. Por isto, realizar uma exploração descritiva de cada situação de acordo com seu polo, seguido da criação de índices de violação e coerência, realizando correlações com variáveis externas. Conseguir chegar a esta resposta contribui com indicações de como instrumentos de autorrelato podem ser construídos e de que maneira erros sistemáticos podem ser mais bem controlados *a priori*.

A segunda hipótese visa se aprofundar nos dois estudos exploratórios anteriores. Ou seja, se as VAs são capazes de captar características sobre o uso da escala da resposta, então o quanto isso estaria relacionado com erros do tipo sistemático, em especial, estilos de resposta? Neste sentido, explorar índices de aquiescência e de viés de grupo e relacioná-los com variáveis externas podem nos indicar as possíveis relações e seus graus de associações.

Método

Participantes

Esta tese consiste na mesma amostra que se encontra o estudo psicométrico inicial para a versão do SENNA 2.0, composta por alunos da rede pública do estado do Ceará, Brasil (N = 11.249). A caracterização desta amostra se deu da seguinte maneira, a saber, 52,4% do sexo feminino, média de idade de 15,2 anos (DP = 2,13). Os anos escolares compreendidos foram o 5º ano (15,5%; faixa etária de 10 a 12 anos) e o 9º ano (43,2%; faixa etária de 13 a 16 anos) do Ensino Fundamental e 1º ano do ensino Médio (41,2% faixa etária de 14 a 17 anos). Esses alunos vieram de 436 escolas localizadas em 110 cidades, e cada escola foi representada por uma média de N = 25 alunos (min = 1, máx = 101).

Instrumentos

Avaliação das Competências Socioemocionais (SENNA v2.0) – Primi et al., 2016b, 2021; Santos & Primi, 2014

O Senna v2.0 é um inventário desenvolvido especificamente para crianças e adolescentes (Primi et al., 2016b; Santos & Primi, 2014) para avaliar as qualidades essenciais das competências socioemocionais (SEMS) usando a taxonomia de personalidade Big Five como uma estrutura abrangente (John et al., 2008; Soto & John, 2018). Uma primeira versão do SENNA foi elaborada a partir de uma revisão sistemática sobre instrumentos socioemocionais até então encontrados na literatura (Santos & Primi, 2014). Um estudo feito com essa primeira versão levou à escolha de um conjunto de 54 itens que melhor apresentaram evidências de validade baseada na estrutura interna e nas relações com variáveis externas.

Nesta coleta, este conjunto de 54 itens foram utilizados, sendo eles distribuídos da seguinte forma: nove itens para a dimensão Autogestão (C; $\alpha = 0,91$), oito itens para a dimensão Abertura à novas ideias (O; $\alpha = 0,78$), dez itens para a dimensão Resiliência Emocional (ER; $\alpha = 0,89$), nove itens para a dimensão Engajamento com os outros (E; $\alpha = 0,83$), dez itens para a

dimensão Amabilidade (A; $\alpha = 0,81$). Ainda, oito itens para a dimensão Locus de Controle (LC; $\alpha = 0,75$). Os itens selecionados para este estudo foram provenientes do estudo de larga escala para a versão final do instrumento Senna v2.0, de forma que informações mais precisas podem ser encontradas no seu manual técnico (Primi et al., 2021). O formato de resposta foi uma escala de classificação Likert com cinco pontos, de forma que 1 era “Nada a ver comigo” e 5 “Muito a ver comigo”. Ainda, constou no Senna um questionário sociodemográfico para caracterização da amostra.

Vinhetas-âncora – Primi et al., 2016b

Cinco conjuntos de vinhetas foram escritos, um para cada dimensão abrangente do SENNA v2.0 (C, O, N, E e A). Cada conjunto de vinhetas retrata descrições curtas de três pessoas que são baixas, médias ou altas em uma dimensão central específico de SEMS, seguidas pelo mesmo item de característica que avalia o domínio correspondente. Os estudantes tiveram que avaliar a posição de cada pessoa hipotética na característica correspondente usando a mesma escala Likert de 5 pontos que os itens de autorrelato do SENNA. Os itens usados para graduar as vinhetas pediam aos alunos que classificassem cada pessoa hipotética em termos de como ela era *Organizada* (para as 3 vinhetas C), *Criativa* (para as 3 vinhetas O), *Nervosa/Preocupada* (para as 3 vinhetas N), *Quanto consegue bater um papo com uma pessoa desconhecida?* (para as 3 vinhetas E) e *Agradável* (para as 3 vinhetas A).

Ainda, cinco conjunto de Vinhetas-âncora com polo negativo para cada dimensão foi escrita, usando os mesmos personagens e histórias acima descritas, mas agora perguntando-se sobre as pessoas usando o polo negativo de cada dimensão. Assim, os itens usados para graduar as vinhetas de polo negativo pediam aos alunos que classificassem cada pessoa hipotética em termos de como ela era *Desorganizada* (para as 3 vinhetas C), *Não interessada em criar coisas novas* (para as 3 vinhetas O), *Insegura* (para as 3 vinhetas N), *Tímida* (para as 3 vinhetas E) e *Egoísta* (para as 3 vinhetas A). A Tabela 3 traz o dicionário dessas vinhetas.

Tabela 3*Dicionário das Vinhetas-âncora com descrição dos itens e seu domínio, polo e nível*

Código do item	Item	Domínio	Polo do adjetivo	Nível do Personagem da vinheta
sv008	Amanda tem muita imaginação, gosta de saber o porquê das coisas. Gosta de inventar histórias e desenhar. Quanto você acha que Amanda é uma pessoa tradicional, que não gosta de coisas muito diferentes?	O	0	Alto
sv010	Jorge é interessado em vários assuntos e gosta de aprender, mas prefere ter uma rotina nos seus trabalhos. Quanto você acha que Jorge é uma pessoa tradicional, que não gosta de coisas muito diferentes?	O	0	Médio
sv012	Luciana tem pouco interesse em arte e música. É mais prática e não gosta de ficar pensando muito sobre um assunto. Quanto você acha que Luciana é uma pessoa tradicional, que não gosta de coisas muito diferentes?	O	0	Baixo
sv011	Luciana tem pouco interesse em arte e música. É mais prática e não gosta de ficar pensando muito sobre um assunto. Quanto você acha que Luciana é criativa?	O	1	Baixo
sv009	Jorge é interessado em vários assuntos e gosta de aprender, mas prefere ter uma rotina nos seus trabalhos. Quanto você acha que Jorge é criativo?	O	1	Médio
sv007	Amanda tem muita imaginação, gosta de saber o porquê das coisas. Gosta de inventar histórias e desenhar. Quanto você acha que Amanda é criativa?	O	1	Alto
sv006	Juliana é bastante cuidadosa e dedicada. Gosta de limpar a casa, é caprichosa nas lições e sempre as termina antes do prazo. Quanto você acha que Juliana é desleixada, não tem cuidado na hora de fazer as coisas?	C	0	Alto
sv004	Manuela tem noção de organização, mas às vezes deixa seu quarto bagunçado. Tende a completar as lições em cima da hora. Quanto você acha que Manuela é desleixada, não tem cuidado na hora de fazer as coisas?	C	0	Médio
sv002	Aline costuma deixar as coisas todas bagunçadas, odeia limpar a casa e deixa as lições de casa sem completar. Quanto você acha que Aline é desleixada, não tem cuidado na hora de fazer as coisas?	C	0	Baixo
sv001	Aline costuma deixar as coisas todas bagunçadas, odeia limpar a casa e deixa as lições de casa sem completar. Quanto você acha que Aline é organizada?	C	1	Baixo
sv003	Manuela tem noção de organização, mas às vezes deixa seu quarto bagunçado. Tende a completar as lições em cima da hora. Quanto você acha que Manuela é organizada?	C	1	Médio
sv005	Juliana é bastante cuidadosa e dedicada. Gosta de limpar a casa, é caprichosa nas lições e sempre as termina antes do prazo. Quanto você acha que Juliana é organizada?	C	1	Alto
sv017	Paulinho é falante e cheio de energia. Tem facilidade para conversar e fazer novas amizades. Quanto você acha que Paulinho é tímido?	E	0	Alto
sv015	Rodrigo é sociável, mas às vezes prefere ficar na dele. Não é de falar muito. Gosta de jogar videogame com os colegas. Quanto você acha que Rodrigo é tímido?	E	0	Médio
sv013	Tiago é mais quieto e tem dificuldade para conversar com os colegas. Nas festas, tende a ficar isolado. Prefere jogar videogame sozinho. Quanto você acha que Tiago é tímido?	E	0	Baixo
sv014	Tiago é mais quieto e tem dificuldade para conversar com os colegas. Nas festas, tende a ficar isolado. Prefere jogar videogame sozinho.	E	1	Baixo

	Quanto você acha que Tiago consegue bater um papo com uma pessoa desconhecida?			
sv016	Rodrigo é sociável, mas às vezes prefere ficar na dele. Não é de falar muito. Gosta de jogar videogame com os colegas. Quanto você acha que Rodrigo consegue bater um papo com uma pessoa desconhecida?	E	1	Médio
sv018	Paulinho é falante e cheio de energia. Tem facilidade para conversar e fazer novas amizades. Quanto você acha que Paulinho consegue bater um papo com uma pessoa desconhecida?	E	1	Alto
sv020	Paula divide suas coisas com os outros, está sempre disposta a ajudar e se preocupa com o que acontece com seus colegas. Quanto você acha que Paula é briguenta e consegue fazer com que outras pessoas façam o que ela quer?	A	0	Alto
sv022	Caio tem consideração pelos outros e é cooperativo, mas às vezes tende a encontrar defeito nos outros. Quanto você acha que Caio é briguento e consegue fazer com que outras pessoas façam o que ele quer?	A	0	Médio
sv024	Julia sempre pensa nela mesma em primeiro lugar, é ciumenta e frequentemente bate-boca com as pessoas. Quanto você acha que Julia é briguenta e consegue fazer com que outras pessoas façam o que ela quer?	A	0	Baixo
sv023	Julia sempre pensa nela mesma em primeiro lugar, é ciumenta e frequentemente bate-boca com as pessoas. Quanto você acha que Julia é amável?	A	1	Baixo
sv021	Caio tem consideração pelos outros e é cooperativo, mas às vezes tende a encontrar defeito nos outros. Quanto você acha que Caio é amável?	A	1	Médio
sv019	Paula divide suas coisas com os outros, está sempre disposta a ajudar e se preocupa com o que acontece com seus colegas. Quanto você acha que Paula é amável?	A	1	Alto
sv029	Pedro é calmo e lida bem com situações tensas e estressantes. Dificilmente se sente triste. Quanto você acha que Pedro é nervoso?	N	0	Alto
sv027	Fabiana lida bem com estresse e confia nas suas capacidades, mas às vezes fica triste e ansiosa. Quanto você acha que Fabiana é nervosa?	N	0	Médio
sv025	Beto se irrita e fica mal-humorado com facilidade. Está sempre preocupado com tudo e tem dificuldade em tomar decisões. Quanto você acha que Beto é nervoso?	N	0	Baixo
sv026	Beto se irrita e fica mal-humorado com facilidade. Está sempre preocupado com tudo e tem dificuldade em tomar decisões. Quanto você acha que Beto é calmo e controla bem seu estresse?	N	1	Baixo
sv028	Fabiana lida bem com estresse e confia nas suas capacidades, mas às vezes fica triste e ansiosa. Quanto você acha que Fabiana é calma e controla bem seu estresse?	N	1	Médio
sv030	Pedro é calmo e lida bem com situações tensas e estressantes. Dificilmente se sente triste. Quanto você acha que Pedro é calmo e controla bem seu estresse?	N	1	Alto

Nota. O – Abertura à novas ideias; C – Conscienciosidade/Autogestão; E – Extroversão/Engajamento com os outros; A – Amabilidade; N – Neuroticismo/Resiliência emocional; 0 – polo negativo; 1 – polo positivo.

Por questões de tempo para aplicação, os alunos responderam apenas seis conjuntos de Vinhetas-âncora, totalizando 18 itens. Assim, na confecção dos cadernos para a coleta, os questionários das VA para as dimensões C e N foram utilizados em todos os cadernos de aplicação, e as demais dimensões foram distribuídos de forma aleatória para todos os cadernos

de aplicação. Desta forma, três versões dos cadernos de VA foram utilizadas, a saber, um com vinhetas para as dimensões C, N, O, um para C, N, E, e o último C, N, A, todos tendo os polos positivos e negativos de cada dimensão.

Medidas de desempenho padronizado (Conhecimento em Língua Portuguesa e Matemática).

Como parte de suas atividades escolares regulares, os alunos concluíram os testes de desempenho padrão para Linguagens (ou seja, português/redação) e Matemática. Estas pontuações para o semestre acadêmico de 2014/01 foram obtidas diretamente dos registros oficiais da secretaria de educação do estado. Vale ressaltar que estas notas padronizadas fazem parte de um programa conhecido como Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAECE, <https://spaece.caedufjf.net>).

Procedimentos

Os alunos responderam aos itens das Vinhetas-âncora, seguido dos itens do SENNA v1. 54. Eles também responderam a um questionário demográfico, tendo sexo, idade, etnia, contexto domiciliar (indicadores econômicos permanentes, que podem ser encontrados em <https://www.abep.org/criterio-brasil>) e comportamentos parentais. A coleta de dados foi realizada durante as atividades regulares em sala de aula em um tempo de aproximadamente 50 minutos. Uma empresa especializada em aplicações de testes contratada pela secretaria do Estado do Ceará para realizar o SPAECE executou toda a coleta de dados. No âmbito dessa coleta a empresa realizou treinamento para que pessoas fora do âmbito escolar, ou seja, que não fossem os professores da escola, realizassem a aplicação destes questionários. A autorização para a realização da coleta ocorreu em parceria com o Instituto Ayrton SENNA e a Secretaria Estadual de Educação do Ceará, em que a presente secretaria permitiu que a coleta fosse realizada em toda a rede estadual, responsabilizando-se pelos alunos. Ainda, a coleta faz parte de um projeto maior, fruto desta parceria, em que intervenções e programas de capacitação

docente seriam elaborados a partir dos resultados oriundos desta avaliação. Esse projeto amplo foi examinado e aprovado pelo comitê de ética em pesquisa da Universidade São Francisco.

Análise de dados

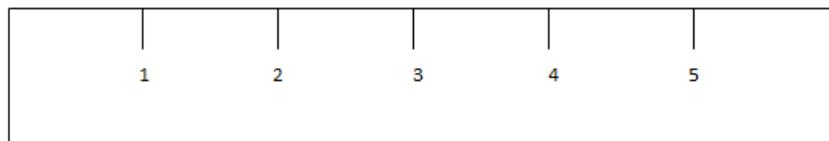
As análises dos dados aconteceram de acordo com as hipóteses testadas. Assim, para verificar a primeira hipótese, isto é, se as Vinhetas-âncora captam diferenças individuais em estilos de resposta, realizou-se um estudo exploratório de caráter descritivo dos dados obtidos pelas respostas as Vinhetas-âncora para verificação da distribuição das respostas em cada um dos polos. Isto possibilitou a criação de índices gráficos para visualizar a maneira como a distribuição ocorre nas situações hipotéticas descritas pelas VAs, bem como o cálculo dos índices de violações do uso das Vinhetas-âncora (King et al., 2010) e um índice global de consistência. Para isto, os pacotes *dplyr* (Wickham et al., 2021), *ggplot2* (Wickham, 2021) e *psych* (Revelle, 2021) foram utilizados.

Os índices de violação das VAs foram calculados a partir do modelo não-paramétrico explicitado por King et al. (2004) e Kyllonen e Bertling (2014) e explicitado na Figura 1. Assim, quanto maior a quantidade de respostas fora do que era esperado, maior esse índice de violação. De forma distinta, o índice global de consistência foi criado para esse estudo da seguinte maneira: o conjunto de resposta dada por cada indivíduo foi correlacionada com o gabarito esperado de cada conjunto das VAs. A Figura 2 representa como as situações combinando o traço da pessoa (nível alto e nível baixo em Amabilidade) e o polo dos adjetivos (positivo e negativo) devem ser endossados caso o sujeito compreenda corretamente o sentido do traço.

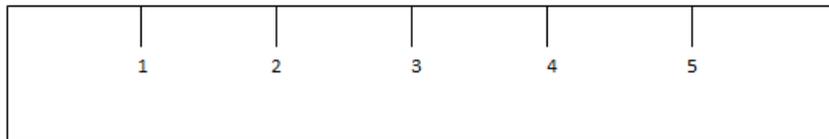
Figura 2

Representação hipotética de endosso das Vinhetas-âncora níveis do traço e o resultado esperado pelo endosso a cada situação.

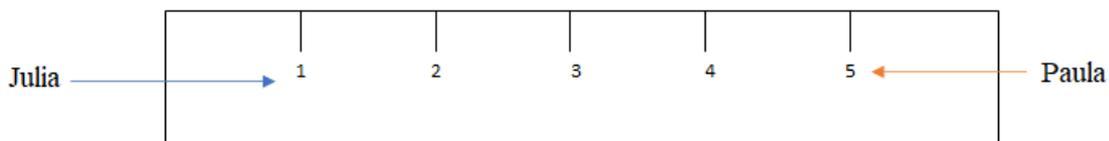
Situação: Paula divide suas coisas com os outros, está sempre disposta a ajudar e se preocupa com o que acontece com seus colegas. (Pessoa com nível alto em Amabilidade).



Situação: Julia sempre pensa nela mesma em primeiro lugar, é ciumenta e frequentemente bate boca com as pessoas. (Pessoa com nível baixo em Amabilidade).



Pergunta: O quanto você acha que (inserir nome do personagem) é amável?



Para investigar evidências de validade externa sobre o uso da escala de resposta das VAs, uma correlação r de Pearson foi feita com cada situação das VAs e o desempenho escolar (isto é, Linguagens e Matemática), variáveis sociodemográficas, mais especificamente série escolar e sexo, e com as dimensões socioemocionais avaliados pela SENNA. Vale salientar que o sexo foi codificado de forma dicotômica, sendo 0 para meninos e 1 para meninas. Em seguida, realizou-se novamente a correlação de Pearson entre os índices de violação das VAs e o índice global de consistência, bem como com o desempenho escolar.

Para a hipótese de que a aquiescência pode ser mensurada pela VAs, isto é, de que as pessoas endossam o item sem se ater ao conteúdo, alguns índices foram calculados para que se pudesse investigar o grau e a direção da associação destes índices com as demais variáveis por meio da correlação r de Pearson. Devido à inclusão de adjetivos com polos negativos e positivos com personagens de níveis baixos e altos no traço, dois tipos de índice de aquiescência podem ser calculados: o primeiro é o índice de aquiescência baseado no traço do personagem, no qual selecionamos pessoas com o mesmo nível baixo podendo ser alto no traço (eliminamos as vinhetas descrevendo pessoas com valor médio no traço) ou baixo e verificamos as respostas em traços opostos para essas pessoas calculando-se a média nesses dois adjetivos de polos opostos. Esse índice indica quanto a pessoa concorda (aquiescência) ou discorda (disaquiescência) dos adjetivos teoricamente opostos, calculando, portanto, respostas inconsistentes semelhantes ao índice de aquiescência tradicional.

O segundo são os índices de aquiescência baseados no atributo avaliado nos personagens, no qual selecionamos um adjetivo indicando um determinado nível do traço (e.g., desorganizada, para um nível baixo de Amabilidade) e verificamos as respostas para personagens das vinhetas com características opostas, novamente alto e baixo (i.e., nível alto em Amabilidade e nível baixo em Amabilidade). Assim esse índice verifica se a pessoa concorda (aquiescência) ou discorda (disaquiescência) que um mesmo adjetivo caracterize pessoas que teoricamente são caracterizadas polos opostos daquela característica. O nome do índice indica qual aspecto foi agregado para seu cálculo – se o traço ou se o atributo. Com estes índices criados, novamente a correlação de Pearson investigou a associação presente entre os índices.

Por fim, para testar a hipótese de que as VAs poderiam mensurar o viés de grupo (GRB), outros índices foram criados. A diferença nestes índices se baseia no quanto a média e o desvio-padrão de um conjunto de Vinhetas-âncora, a partir do seu conteúdo, pode variar ou não. Assim,

uma média para cada conjunto de Vinhetas-âncora foi calculada a partir das respostas dadas, junto com o desvio-padrão. Note que diferentemente dos índices de aquiescência, que usou apenas pares semânticos ou níveis distintos, aqui se utilizou todas as respostas de cada conjunto. Para os conjuntos de VAs de polo negativo, as respostas foram invertidas antes de se criar esses índices, a fim de que um escore mais alto e mais fidedigno para os índices fossem gerados. Por fim, um índice geral da média e desvio-padrão para cada o conjunto de polo positivo e para o polo negativo foram criados.

Desta forma, com estes índices, pode-se verificar a distribuição das frequências para os índices gerais, por meio de uma análise descritiva por meio de histogramas. Também correlações de Pearson destes índices com as SEMS, notas do desempenho padrão em Linguagens e Matemática e com as variáveis sociodemográficas sexo e série escolar podem nos ajudar a testar a validade das VAs para GRB.

Resultados

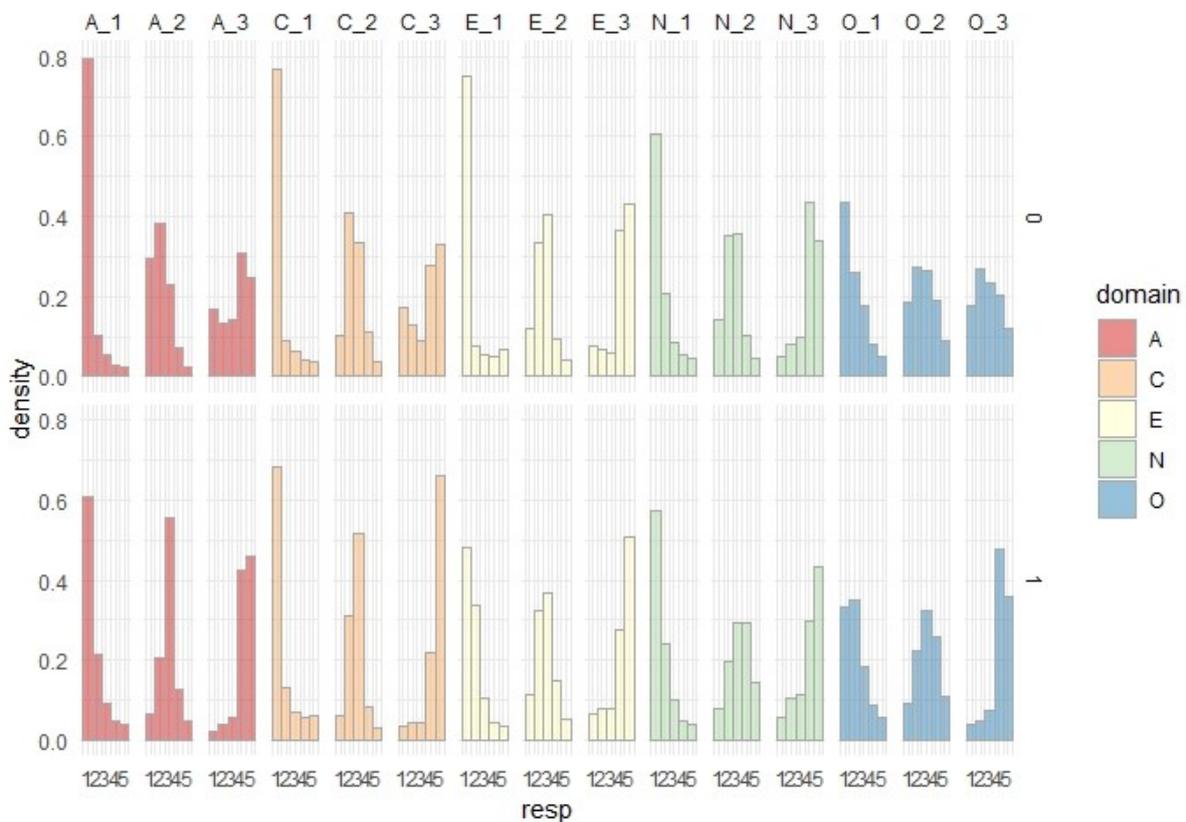
O objetivo deste estudo foi verificar as evidências de validade para mensuração de estilos de respostas por meio da técnica conhecida como Vinhetas-âncora (VAs). Os resultados a seguir são apresentados de acordo com as hipóteses testadas e as análises realizadas.

Análises descritivas das Vinhetas-âncora

Para averiguar as características descritivas do uso das VAs, uma distribuição geral das respostas para cada item em cada um dos conjuntos foi calculada. Este resultado visa testar a primeira hipótese, qual seja, se as Vinhetas-âncora são capazes de ser um instrumento que captam as diferenças individuais dos respondentes no tocante ao erro sistemático, isto é, estilos de resposta. Isto se torna possível devido ao uso do conjunto de VAs de polo distintos para o mesmo domínio, em que se pode realizar a inversão dos escores endossados pela mesma pessoa em situações idênticas, mas com o polo distinto. A Figura 3 apresenta estes resultados gráficos, sendo uma média das frequências das respostas obtidas em cada nível do traço latente avaliado.

Figura 3.

Histograma com as distribuições das respostas de cada conjunto das Vinhetas-âncora.



Nota. A: Agradabilidade; C: Autogestão. E: Engajamento com os outros. N: Resiliência emocional. O: Abertura à novas ideias. 0: conjunto de vinhetas negativos. 1: conjunto de vinhetas positivos. “_1” : situação 1, com baixo nível no traço latente. “_2” : situação 2, com nível médio do traço latente. “_3” : situação 3, com alto nível no traço latente.

No que tange a verificação das distribuições das respostas, a Figura 3 demonstra como cada conjunto de vinhetas teve sua distribuição em cada uma das situações, tanto para o conjunto de vinhetas de polo negativo quanto de polo positivo. Inicia-se descrevendo os resultados verificados do conjunto de Vinhetas-âncoras de polo negativo. Elas apresentaram uma frequência muito alta de resposta para o nível baixo do traço latente, como esperado, haja vista que nestas situações sempre houve um adjetivo representando o polo baixo do traço perguntado sobre o personagem da história. As dimensões Resiliência emocional e Abertura à novas ideias tiveram uma frequência menor quando comparados aos demais domínios, mas ainda assim conseguiram apresentar uma curva assimétrica positiva na sua distribuição, como

era esperado. Já no nível médio do traço latente, que era esperado uma curva simétrica da distribuição, isto não pôde ser observado neste conjunto de Vinhetas-âncora, com exceção do domínio Abertura à novas ideias. Contudo, mesmo que este domínio possa ter apresentado simetria esperada, é possível notar que outras chaves de respostas também aparecem com frequência próxima à distribuição da chave 3, formando uma curva platicúrtica. O nível alto do traço latente, cuja distribuição esperada era semelhante ao nível baixo do traço, obteve uma distribuição semelhante ao nível médio.

Quanto ao conjunto de VAs positivos, os resultados observados são mais heterogêneos que os do conjunto negativo. Para começar, o nível de baixo traço latente apresentou em todos os domínios, a exceção de Abertura `novas ideias, uma frequência nas chaves esperadas igual ou superior a 80%, numa distribuição assimétrica. De forma semelhante, o nível de alto traço latente apresentou para todos os domínios uma frequência igual ou superior a 80%, com uma distribuição assimétrica. Já para o nível médio de traço latente, todos os resultados apresentaram uma curva simétrica. Contudo, apenas as dimensões Amabilidade e Autogestão apresentaram uma curva de distribuição leptocúrtica, indicando que a chave de resposta 3 foi a mais frequente dos casos, enquanto os demais domínios apresentaram uma curva mesocúrtica.

Visando avançar na compreensão sobre a distribuição das respostas, e suas frequências, nos conjuntos de polo positivo e polo negativo, uma distribuição geral foi calculada de forma a apresentar, de uma maneira visual, os padrões de respostas de acordo com a polaridade dos conjuntos e as respostas esperadas. A Figura 4 apresenta estes resultados.

Figura 4

Histograma com as distribuições das médias de frequência das respostas nas Vinhetas-âncora em todos os domínios



Nota. Para ‘Order’ 1: itens que a resposta esperada seja 1; 2: itens que a resposta esperada seja 2; 3: itens que resposta esperada seja 3. No eixo vertical à direita, 0: conjunto Vinhetas-âncora de polo negativo; 1: conjunto Vinhetas-âncora de polo positivo.

Aqui se observa, em linhas gerais, que as VAs de itens de polo negativo apresentaram uma curva de distribuição assimétrica e leptocúrtica para o nível alto de traço latente. Contudo, as demais distribuições deste conjunto não demonstraram uma prevalência na frequência esperada. Como se pode observar, o nível médio do traço latente apresentou uma certa simetria, mas esta curva teve uma característica de predominância mesocúrtica, como o que aconteceu com o nível baixo de traço latente, ainda que este apresentasse uma distribuição assimétrica.

Na contramão dos resultados achados do conjunto de vinhetas de polo negativo, as VAs de polo positivo, na média de frequência das respostas, apresentaram curvas de distribuição

dentro do esperado. Isto é, os níveis de baixo e alto de traço latente tiveram uma curva assimétrica e leptocúrtica, indicando que as chaves extremas foram utilizadas em maior quantidade nestes itens e de maneira correta, enquanto o nível médio de traço latente teve uma curva simétrica, indicando uma tendência a resposta para a chave 3. Nota-se que este nível de traço latente também apresentou uma frequência de resposta considerável para a chave 2, tornando uma curva de distribuição mesocúrtica.

Validade externa dos escores das Vinhetas-âncora

Para investigar a relação entre traços pessoais e o endosso às situações expostas pelas Vinhetas-âncora, utilizou-se a correlação de Pearson para aferir o grau e a direção dessa relação com algumas variáveis externas. Dentre elas, estão as notas em Linguagens e Matemática, as variáveis socioemocionais avaliados pelo SENNA, bem como gênero e série escolar. A Tabela 4 apresenta os resultados da correlação entre as VAs e os notas de Linguagens e Matemática.

Tabela 4*Correlação das situações das Vinhetas-âncora com as notas escolares*

Código da Vinheta-âncora	Linguagens	Matemática
sv001_C_lw_Aline_organizada	-0.12**	-0.12**
sv002_C_lw_Aline_desleixada	0.16**	0.15**
sv003_C_av_Manuela_organizada	0.05**	0.06**
sv004_C_av_Manuela_desleixada	0.02	0.01
sv005_C_hi_Juliana_organizada	0.21**	0.19**
sv006_C_hi_Juliana_desleixada	-0.22**	-0.18**
sv007_O_hi_Amanda_criativa	0.20**	0.19**
sv008_O_hi_Amanda_tradicional	-0.20**	-0.14**
sv009_O_av_Jorge_criativa	-0.12**	-0.08**
sv010_O_av_Jorge_tradicional	0.12**	0.10**
sv011_O_lw_Luciana_criativa	-0.16**	-0.13**
sv012_O_lw_Luciana_tradicional	0.13**	0.12**
sv013_E_lw_Tiago_tímido	0.21**	0.17**
sv014_E_lw_Tiago_desinibido	-0.15**	-0.12**
sv016_E_av_Rodrigo_desinibido	0.09**	0.09**
sv015_E_av_Rodrigo_tímido	-0.01	0.01
sv017_E_hi_Paulinho_tímido	-0.24**	-0.19**
sv018_E_hi_Paulinho_desinibido	0.26**	0.22**
sv019_A_hi_Paula_amável	0.17**	0.13**
sv020_A_hi_Paula_briguenta	-0.24**	-0.17**
sv021_A_av_Caio_amável	0.10**	0.11**
sv022_A_av_Caio_briguenta	-0.07**	-0.03
sv023_A_lw_Julia_amável	-0.21**	-0.12**
sv024_A_lw_Julia_briguenta	0.23**	0.17**
sv025_N_lw_Beto_nervoso	0.24**	0.20
sv026_N_lw_Beto_calmo	-0.26**	-0.19**
sv027_N_av_Fabiana_nervoso	-0.04**	-0.03**
sv028_N_av_Fabiana_calmo	0.14**	0.11**
sv029_N_hi_Pedro_nervoso	-0.25**	-0.18**
sv030_N_hi_Pedro_calmo	0.27**	0.22**

Nota. ** $p \leq 0.01$; C – Autogestão; E – Engajamento com os outros; N – Resiliência emocional; A – Amabilidade; O – Abertura à novas ideias; NVloc – Locus de controle negativo; ‘sv’ – Situação da Vinheta-âncora; ‘hi’ – Situação alto no traço; ‘av’ – Situação de traço médio; ‘lw’ – Situação de traço baixo.

Os resultados apresentados na Tabela 4 demonstraram que as situações expostas das VAs de nível alto e de nível baixo apresentam associações estatisticamente significativas, sendo de direção positivo no caso das VAs do conjunto de polo positivo, e direção negativa quando do conjunto de polo negativo. Percebe-se que o grau das associações para os atributos avaliados nos níveis alto e baixo possuem pequenas diferenças, de modo que as situações dos conjuntos

de polo positivo tem um grau um pouco maior em detrimento às situações dos conjuntos de polo negativo. Por sua vez, as situações de nível médio variaram em termos de significância estatística, e o grau da associação foi mais tênue. A direção, contudo, assemelhou-se às situações dos outros níveis.

Em seguida, explorando ainda mais a associação de como cada item pode estar relacionado a um dos traços de personalidade de forma mais específica, cada situação apresentada pelas Vinhetas-âncora foi correlacionada com as dimensões das competências socioemocionais (SEMS) do SENNA. A Tabela 5 apresenta estes resultados.

Tabela 5*Correlação das situações das Vinhetas-âncora com as dimensões do SENNA*

Código da Vinheta-âncora	C	E	N	A	O	NVLoc
sv001_C_lw_Aline_organizada	-0.08**	-0.04**	-0.04**	-0.06**	-0.03	0.09**
sv002_C_lw_Aline_desleixada	0.04	0.06**	0.03	0.10**	0.09**	-0.04**
sv003_C_av_Manuela_organizada	-0.02	0.03*	0.00	0.03	0.04**	0.04**
sv004_C_av_Manuela_desleixada	0.00	0.01	-0.01	0.02	0.04**	0.02
sv005_C_hi_Juliana_organizada	0.14**	0.14**	0.07**	0.18**	0.12**	-0.10
sv006_C_hi_Juliana_desleixada	-0.08**	-0.09**	-0.06**	-0.12**	-0.05**	0.13**
sv007_O_hi_Amanda_criativa	0.13**	0.15**	0.08**	0.22**	0.20**	-0.05**
sv008_O_hi_Amanda_tradicional	-0.03	-0.07**	-0.04	-0.07**	-0.03	0.13**
sv009_O_av_Jorge_criativa	0.11**	0.02	0.04	0.06**	0.06**	0.07**
sv010_O_av_Jorge_tradicional	-0.01	0.02	-0.03	0.02	0.10**	0.02
sv011_O_lw_Luciana_criativa	0.00	-0.04	-0.04	-0.04	-0.01	0.12**
sv012_O_lw_Luciana_tradicional	-0.01	0.05*	0.00	0.08**	0.09**	-0.01
sv013_E_lw_Tiago_tímido	0.08**	0.12**	0.03	0.16**	0.12**	-0.05**
sv014_E_lw_Tiago_desinibido	-0.08	-0.06**	-0.04*	-0.09**	-0.04	0.10**
sv015_E_av_Rodrigo_tímido	0.01	-0.01	-0.03	0.02	0.03	0.06**
sv016_E_av_Rodrigo_desinibido	-0.03	0.05**	0.02	0.02	0.07**	0.00
sv017_E_hi_Paulinho_tímido	-0.04**	-0.09**	-0.05**	-0.11**	-0.06**	0.13**
sv018_E_hi_Paulinho_desinibido	0.05**	0.18**	0.06**	0.17**	0.15**	-0.10**
sv019_A_hi_Paula_amável	0.14**	0.21**	0.07*	0.33**	0.17**	-0.09**
sv020_A_hi_Paula_briguenta	-0.11**	-0.10**	-0.10**	-0.18**	-0.06	0.18**
sv021_A_av_Caio_amável	0.05	0.04	-0.01	0.08**	0.09**	0.00
sv022_A_av_Caio_briguenta	-0.09**	0.00	-0.07**	-0.09**	0.01	0.10**
sv023_A_lw_Julia_amável	-0.04	-0.09**	-0.11**	-0.16**	-0.04	0.15**
sv024_A_lw_Julia_briguenta	0.02	0.12**	0.04	0.13**	0.11**	-0.07**
sv025_N_lw_Beto_nervoso	0.03	0.12**	-0.01	0.14**	0.13**	-0.03
sv026_N_lw_Beto_calmo	-0.06**	-0.11**	-0.05*	-0.14**	-0.04	0.18**
sv027_N_av_Fabiana_nervoso	0.01**	0.00	-0.06**	0.00	0.04	0.07*
sv028_N_av_Fabiana_calmo	0.07**	0.09**	0.08**	0.12**	0.09**	-0.05*
sv029_N_hi_Pedro_nervoso	-0.05*	-0.09**	-0.07**	-0.11**	-0.05*	0.15**
sv030_N_hi_Pedro_calmo	0.11**	0.14**	0.09**	0.19**	0.17**	-0.10**

Nota. ** $p \leq 0.01$; C – Autogestão; E – Engajamento com os outros; N – Resiliência emocional; A – Amabilidade; O – Abertura à novas ideias; NVloc – Locus de controle negativo; ‘sv’ – situação da Vinheta-âncora; ‘hi’ – Situação de traço alto; ‘av’ – Situação de traço médio; ‘lw’ – Situação de traço baixo.

Os resultados apresentados na Tabela 5 trouxeram informações importantes sobre a relação no desempenho da tarefa com traços das competências socioemocionais avaliados pelo SENNA. Percebe-se que os padrões de correlações encontrados tendem a ser semelhantes com os padrões de correlações encontrados com as notas de Linguagens e Matemática, em especial a dimensão de Amabilidade. Um ponto importante é que as direções negativas das associações,

com exceção da dimensão de Locus de Controle, aconteceram para o conjunto de vinhetas de polo negativo. Desta forma, os estudantes que compreendem melhor o sentido das vinhetas, refletem em um endosso com maior congruência e tendem a apresentar escores maiores nas capacidades socioemocionais nas dimensões do SENNA.

No que tange às características sociodemográficas, as situações apresentadas pelas Vinhetas-âncora foram correlacionadas com o gênero e o nível escolar. Salienta-se que o sexo masculino foi codificado como '0', enquanto o feminino foi codificado como '1' – essa informação é relevante para compreender sobre a direção das associações. Por último, foi realizado uma correlação com o nível escolar geral de toda a amostra e, em seguida, com os anos escolares de formas separadas. A Tabela 6 apresenta estes resultados.

Tabela 6*Correlação das situações das Vinhetas-âncora com Sexo e Série escolar*

	Sexo	Série geral	EF_5	EF_9	EM_1
sv001_C_lw_Aline_organizada	-0,04**	0,02	-0,02	0,02	0,00
sv002_C_lw_Aline_desleixada	0,03	-0,05**	0,03	0,04**	-0,06**
sv003_C_av_Manuela_organizada	-0,04**	-0,01	0,00	0,02	-0,02
sv004_C_av_Manuela_desleixada	-0,02	0,00	0,00	-0,01	0,01
sv005_C_hi_Juliana_organizada	0,07**	0,00	0,00	0,02	-0,02
sv006_C_hi_Juliana_desleixada	-0,09**	0,01	-0,01	-0,02	0,03
sv007_O_hi_Amanda_criativa	0,04	-0,04	0,04	0,00	-0,03
sv008_O_hi_Amanda_tradicional	-0,06**	-0,04	0,03	0,00	-0,03
sv009_O_av_Jorge_criativa	0,01	-0,11	0,12**	-0,06**	-0,03**
sv010_O_av_Jorge_tradicional	-0,02	0,08**	-0,09**	0,05**	0,02
sv011_O_lw_Luciana_criativa	-0,04	-0,08	0,09**	-0,05**	-0,02*
sv012_O_lw_Luciana_tradicional	0,01	0,03	-0,04	0,05	-0,01
sv013_E_lw_Tiago_tímido	0,06	0,02**	-0,03	0,06**	-0,04
sv014_E_lw_Tiago_desinibido	-0,09**	0,03	-0,03	0,01	0,02
sv015_E_av_Rodrigo_tímido	-0,02	0,00	0,00	0,00	0,00
sv016_E_av_Rodrigo_desinibido	-0,06**	0,06**	-0,06**	0,03	0,02
sv017_E_hi_Paulinho_tímido	-0,08	-0,07**	0,09**	-0,07**	0,01
sv018_E_hi_Paulinho_desinibido	0,04	0,10**	-0,11**	0,07**	0,01
sv019_A_hi_Paula_amável	0,10**	-0,02	0,02	0,02	-0,03
sv020_A_hi_Paula_briguenta	-0,10	-0,01**	0,03	-0,06	0,04
sv021_A_av_Caio_amável	0,03	-0,02	0,03	-0,04	0,02
sv022_A_av_Caio_briguenta	-0,04	0,04	-0,04	0,03	0,00
sv023_A_lw_Julia_amável	-0,05	-0,06	0,07*	-0,06	0,01
sv024_A_lw_Julia_briguenta	0,04	0,08	-0,09**	0,04**	0,02
sv025_N_lw_Beto_nervoso	0,04	0,02	-0,04	0,06	-0,04**
sv026_N_lw_Beto_calmo	-0,09**	-0,03	0,04	-0,05	0,02*
sv027_N_av_Fabiana_nervoso	0,02	0,01	-0,01	0,01	0,00
sv028_N_av_Fabiana_calmo	0,02	0,02	-0,02	0,01	0,00
sv029_N_hi_Pedro_nervoso	-0,07	-0,07**	0,08**	-0,04**	-0,01
sv030_N_hi_Pedro_calmo	0,04	0,09	-0,10**	0,06**	0,01**

Nota. * $p < 0,05$; ** $p < 0,01$; Série geral: alunos do 5º e 9º ano do Fundamental I e 1º do Ensino Médio; EF_5 – alunos do 5º ano do Fundamental I; EF_9 – alunos do 9º ano do Fundamental I; EM_1 – 1º ano do Ensino Médio; C – Autogestão; E – Engajamento com os outros; N – Resiliência emocional; A – Amabilidade; O – Abertura à novas ideias; NVloc – Locus de controle negativo; ‘sv’ – situação da Vinheta-âncora; ‘hi’ – Situação de traço alto; ‘av’ – Situação de traço médio; ‘lw’ – Situação de traço baixo.

Como demonstrado na Tabela 6, as associações significativas entre gênero e Vinhetas-âncora foram observadas principalmente para as situações de Autogestão e em alguns casos para as situações descritas nas vinhetas para Engajamento com os outros e um caso descrito para as dimensões Amabilidade e Resiliência Emocional. O sentido das correlações indicou que

as meninas tenderam a dar respostas mais coerentes que os meninos, condizendo com os achados indicando uma maior habilidade verbal. Em relação ao ano escolar, o que se percebeu é que correlações foram mais fracas e indicaram que alunos mais novos (5º ano) tenderam a apresentar padrões mais inconsistentes. As correlações significativas ocorreram em maior número comparando-se os alunos do 5º ano (EF_5) com os alunos restantes.

Vinhetas como um teste de conhecimento do vocabulário dos testes socioemocionais

Um ponto que merece atenção a ser explorado como evidências de validade das Vinhetas-âncora são as possíveis correlações existentes entre os índices de violação e o índice global de consistência das respostas às VAs entre si e com o desempenho escolar. Como explicado, os índices de violação demonstram o quanto os participantes respondem dentro da maneira esperada (i.e., chaves de resposta mais baixa para uma situação que descreve um nível baixo de traço latente). Isto porque, como demonstrado nas Tabelas 5 e 6, parece haver uma relação entre a compreensão da tarefa e do conteúdo das situações, que remetem às SEMS, com o desempenho nas VAs. Assim, apresenta-se, inicialmente, as correlações encontradas entre o índice global de consistência das Vinhetas-âncora e os índices de violações das dimensões específicas, como consta na Tabela 7.

Tabela 7

Correlação entre o índice global de consistência e os índices de violações das Vinhetas-âncora

	glb_vig_consist	A.1.viol	C.0.viol	C.1.viol	E.0.viol	E.1.viol	N.0.viol	N.1.viol
glb_vig_consist								
A.1.viol	-.53							
C.0.viol	-.53	.13						
C.1.viol	-.47	.23	.12					
E.0.viol	-.59	.25	.23	.19				
E.1.viol	-.50	.29	.21	.18	.25			
N.0.viol	-.56	NA	.21	.18	.37	.22		
N.1.viol	-.54	NA	.21	.19	.28	.26	.29	
O.1.viol	-.47	NA	.17	.18	.24	.17	.19	.20

Nota. glb_vig_consist: Consistência global das Vinhetas-âncora; A: Amabilidade; C: Autogestão; E: Engajamento com os outros; N: Resiliência emocional; O: Abertura à novas ideias; ‘.0.viol’: Escore médio das violações de Vinhetas-âncora de polo negativo; ‘.1.viol’: Escore médio das violações de Vinhetas-âncora de polo negativo; * $p \leq 0.05$; ** $p \leq 0.01$

Destaca-se que, entre si, as correlações entre os índices de violações das Vinhetas-âncora foram positivas, e o grau destas associações variaram entre $r = .12$ e $r = .37$, de forma que os índices de violação da dimensão Engajamento com os Outros tiveram resultados mais expressivos. Em relação ao índice de consistência global, todos os resultados foram de direção negativa e de moderada magnitude. Assim, isso indica que quanto mais consistentes forem as respostas dos sujeitos de acordo com o esperado, menores os índices de violações observados. Esses dados sugerem que as VA são uma tarefa cognitiva de compreensão de texto, sendo que o gabarito de respostas corretas são as respostas esperadas como operacionalizadas no índice de consistência. As correlações entre os itens e a correlação com o índice global de consistência são condizentes no sentido e magnitude com correlações interitem e correlação item total de testes cognitivos.

Se essa interpretação for correta, o índice global de consistência deve se correlacionar positivamente com desempenho (e mais alto com Linguagens) e as violações devem correlacionar negativamente. Assim para checar essa hipótese se a consistência na forma de uso da escala está relacionada com algum aspecto cognitivo, as correlações com o desempenho escolar foram realizadas. Estes resultados estão descritos na Tabela 8.

Tabela 8

Correlação entre o índice de consistência global e os índices de violações com desempenho escolar

	Linguagens	Matemática
glb_vig_consist	.39	.32
A.1.viol	-.22	-.15
C.0.viol	-.19	-.16
C.1.viol	-.17	-.16
E.0.viol	-.21	-.16
E.1.viol	-.20	-.17
N.0.viol	-.24	-.18
N.1.viol	-.26	-.20
O.1.viol	-.20	-.18

Nota. glb_vig_consist: Consistência global das Vinhetas-âncora; A: Amabilidade; C: Autogestão; E: Engajamento com os outros; N: Resiliência emocional; O: Abertura à novas ideias; ‘.0.viol’: Escore médio das violações de Vinhetas-âncora de polo negativo; ‘.1.viol’: Escore médio das violações de Vinhetas-âncora de polo negativo; * $p \leq 0.05$; ** $p \leq 0.01$

Nestes resultados, observa-se que o desempenho escolar e o índice de consistência global se relacionam de maneira positiva e com um grau entre moderado e alto, enquanto os demais índices de violações demonstraram relação de direção negativa corroborando a hipótese levantada. Conforme o esperado, estes resultados indicam que o uso da escala de resposta de um questionário é mais bem compreendido, e, conseqüentemente, utilizado por respondentes com maiores habilidades cognitivas, em especial com melhor capacidade de leitura e compreensão de texto.

Mensuração de estilo de resposta aquiescente por Vinhetas-âncora

Nessa parte do trabalho testamos se é possível medir o estilo de resposta aquiescente a partir das Vinhetas-âncora. Para isso, como descrito na seção ‘Método’, criamos dois índices. O primeiro índice é uma média de concordância em adjetivos de par semântico de polos opostos, que descrevem um mesmo personagem da vinheta. Ou seja, a situação do personagem Pedro, que é calmo e lida bem com estresse, uma média foi criada a partir das respostas a esta vinheta no polo positivo e negativo. Assim, em uma escala de cinco pontos, as respostas

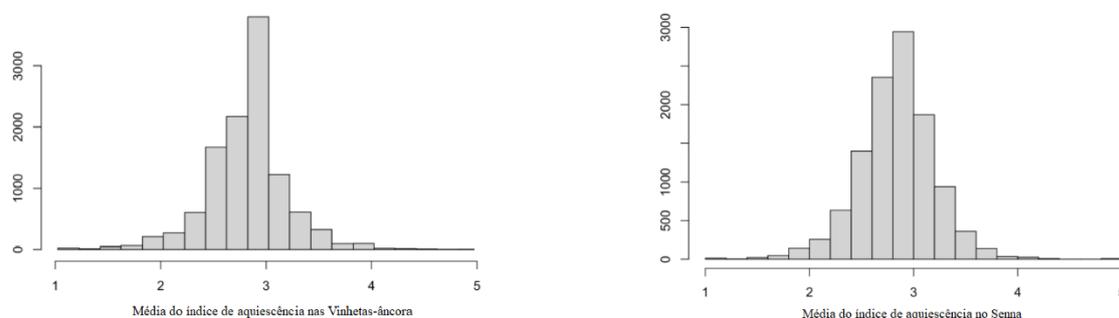
esperadas do conjunto das VAs seriam 12 x 45 com média esperada de três. Nomeamos este índice como um Índice de Aquiescência no Traço (dois adjetivos opostos descrevendo um mesmo traço).

O segundo índice também é uma média, mas leva-se em consideração dois personagens das vinhetas, um exemplificando o nível alto e outro o nível baixo no traço, mas sendo avaliados pelo mesmo adjetivo (e.g.: cuidadoso, na situação que descreve Aline, bagunceira e desorganizada, e Juliana, cuidadosa e dedicada). Portanto, também é esperado escore médio três para este índice. Nomeamos este índice como um Índice de Aquiescência no Atributo (um mesmo adjetivo descrevendo dois personagens com níveis opostos). Finalmente temos um Índice de Aquiescência Tradicional a partir do autorrelato no SENNA pela média de da concordância nos pares semânticos no teste.

Para averiguar a distribuição de como a aquiescência acontece nas Vinhetas-âncora, criamos uma média geral de aquiescência para as VAs e comparamos com a distribuição do índice de aquiescência do SENNA. A Figura 5 mostra o histograma destes índices.

Figura 5

Histograma das médias de aquiescência das Vinhetas-âncora (esquerda) e SENNA (direita).



Na Figura 5 são apresentadas as distribuições dos índices de aquiescência nas vinhetas (média dos dois índices de aquiescência traço e atributo) e no SENNA. Os histogramas são considerados simétricos, denotando uma distribuição normal ao redor de um valor ligeiramente menor do que três, mas dentro do esperado. Isso sugere que, em geral, há mais discordância aos itens do que o esperado, talvez por causa da discordância mais extrema nos itens negativos conforme visualizamos nas vinhetas. Ou seja, mostra uma tendência em concordar ou discordar de maneira consistente com a situação descrita, tanto nas situações expostas pelas VAs, quanto nas situações do SENNA.

A partir da Tabela 9, pode-se perceber que o resultado da correlação entre os índices de aquiescência do SENNA e o índice de aquiescência das Vinhetas-âncora apresentou um grau fraco ($r = 0.21$) e direção positiva.

Na Tabela 9 temos as correlações entre os itens de aquiescência no traço. Apresenta-se o índice de aquiescência geral no traço e os índices de aquiescência no traço para cada dimensão socioemocional avaliada pelo SENNA.

Tabela 9

Correlação entre os índices de aquiescência no traço para cada situação e geral

	A_hi_Paula	A_lw_Julia	C_hi_Juliana	C_lw_Aline	E_hi_Paulinho	E_lw_Tiago	N_hi_Pedro	N_lw_Beto	O_hi_Amanda	O_lw_Luciana
A_lw_Julia	0.14**	1.00								
C_hi_Juliana	0.26**	0.04**	1.00							
C_lw_Aline	0.05**	0.19**	-0.01	1.00						
E_hi_Paulinho	0.14**	0.05**	0.12**	0.04**	1.00					
E_lw_Tiago	0.12**	0.19**	0.12**	0.09**	0.07**	1.00				
N_hi_Pedro	-	-	0.12**	0.07**	0.17**	0.14**	1.00			
N_lw_Beto	-	-	0.11**	0.09**	0.07**	0.22**	0.12**	1.00		
O_hi_Amanda	-	-	0.21**	0.04**	0.07**	0.20**	0.12**	0.15**	1.00	
O_lw_Luciana	-	-	0.09**	0.11**	0.08**	0.12**	0.10**	0.15**	0.09**	1.00
acq_index_trait	0.48**	0.58**	0.44**	0.54**	0.47**	0.51**	0.50**	0.49**	0.52**	0.55**

Nota. ** $p < 0,05$; '-' valor não calculado; A: Amabilidade; C: Autogestão; E: Engajamento com os outros; N: Resiliência emocional; O: Abertura à novas ideias; 'lw' – situação da Vinheta-âncora de traço baixo; 'hi' – situação da Vinheta-âncora de traço alto; acq_index_trait – índice de média de aquiescência no traço

Os resultados das correlações apresentados mostraram que os índices de aquiescência nos atributos tiveram um grau de associação moderado com o Índice de Atributo Geral, conforme era esperado. Entre os graus de associações entre os índices de aquiescência nos traços, esses valores variaram $r = 0,05$ e $r = 0,26$. Destaca-se os resultados do índice de aquiescência ‘E_lw_Tiago’ que apresentou graus fracos com quase todos os demais índices, inclusive apresentando um padrão diferente de associações do seu índice oposto, ‘E_hi_Paulinho’ com as dimensões Resiliência Emocional e Abertura à Novas Ideias.

Por sua vez, a Tabela 10 apresenta as correlações entre os itens de Aquiescência no Atributo. De forma semelhante, aqui temos o Índice de Aquiescência Geral no atributo e os Índices de Aquiescência no Atributo para cada dimensão socioemocional avaliada pelo SENNA

Tabela 10

Correlação entre os índices de aquiescência no atributo para cada situação e geral

	A_0_briguenta	A_1_amável	C_0_desleixada	C_1_organizada	E_0_tímido	E_1_desinibido	N_0_nervoso	N_1_calmo	O_0_tradicional	O_1_criativa
A_1_amável	0.03	1.00								
C_0_desleixada	0.22**	0.04	1.00							
C_1_organizada	0.07**	0.16**	0.05**	1.00						
E_0_tímido	0.09**	0.14**	0.08**	0.11**	1.00					
E_1_desinibido	0.13**	0.07**	0.07**	0.07**	0.00	1.00				
N_0_nervoso	-	-	0.10**	0.10**	0.23**	0.07**	1.00			
N_1_calmo	-	-	0.08**	0.07**	0.06**	0.20**	-0.06**	1.00		
O_0_tradicional	-	-	0.14**	0.04**	0.09**	0.11**	0.11**	0.10**	1.00	
O_1_criativa	-	-	0.06**	0.17**	0.17**	0.06**	0.16**	0.09**	0.01	1.00
acq_index_attrib	0.56**	0.46**	0.53**	0.48**	0.49**	0.46**	0.48**	0.45**	0.53**	0.50**

Nota. ** $p < 0,05$; A: Amabilidade; C: Autogestão; E: Engajamento com os outros; N: Resiliência emocional; O: Abertura à novas ideias; ‘0’ – situação de Vinheta-âncora de polo negativo; ‘1’ – situação de Vinheta-âncora de polo positivo; acq_index_attrib – índice de média de aquiescência no atributo.

Os resultados das correlações entre os índices de aquiescência no atributo para cada situação e geral demonstraram, em suma, associações de direções positivas e graus de associação significativos variando de $r = 0,05$ a $r = 0,23$ para as situações e $r = 0,45$ a $r = 0,56$ para o geral. Um resultado que chama a atenção é a associação negativa entre os índices de

aquiescência de Resiliência emocional (N_0_nervoso e N_1_calmo), que foi a única direção negativa observada nas associações.

Uma terceira correlação de Pearson foi testada para identificar os graus de associação e a direção destas associações entre os diferentes índices de aquiescência no traço e no atributo, ou seja, tanto uma correlação entre os índices das situações quanto os índices gerais. Também é mostrado a correlação dos índices com o desvio-padrão do índice geral de aquiescência no SENNA. A Tabela 11 traz estes resultados.

Tabela 11

Correlação entre os índices de aquiescência nos atributos e nos traços para cada situação e índices gerais de aquiescência

	acq_index_trait	acq_index_attrib	acq_avr	acq_sd
acq_index_trait	1.00			
acq_index_attrib	0.96**	1.00		
acq_avr	0.21**	0.21**	1.00	
acq_sd	0.08**	0.08**	0.21**	1.00
A_hi_Paula	0.48**	0.47**	0.11**	0.07**
A_lw_Julia	0.58**	0.57**	0.11**	-0.03
C_hi_Juliana	0.44**	0.43**	0.08**	0.05**
C_lw_Aline	0.54**	0.53**	0.08**	0.02*
E_hi_Paulinho	0.47**	0.46**	0.09**	0.06**
E_lw_Tiago	0.51**	0.50**	0.13**	0.03**
N_hi_Pedro	0.50**	0.49**	0.16**	0.06**
N_lw_Beto	0.49**	0.49**	0.14**	0.03*
O_hi_Amanda	0.52**	0.51**	0.16**	0.08**
O_lw_Luciana	0.55**	0.54**	0.12**	0.01
A_0_briguenta	0.56**	0.56**	0.09**	-0.03*
A_1_amável	0.44**	0.46**	0.13**	0.06**
C_0_desleixada	0.52**	0.53**	0.07**	0.02**
C_1_organizada	0.47**	0.48**	0.09**	0.04**
E_0_tímido	0.47**	0.49**	0.10**	0.07**
E_1_desinibido	0.45**	0.46**	0.11**	0.03*
N_0_nervoso	0.46**	0.48**	0.14**	0.04**
N_1_calmo	0.44**	0.45**	0.14**	0.05**
O_0_tradicional	0.52**	0.53**	0.13**	-0.02
O_1_criativa	0.49**	0.50**	0.14**	0.10**

Nota. ** $p < 0,01$; acq_index_trait – índice de aquiescência no traço; acq_index_attrib – índice de aquiescência no atributo; acq_avr – média do índice de aquiescência; acq_sd – desvio-padrão do índice de aquiescência; A: Amabilidade; C: Autogestão; E: Engajamento com os outros; N: Resiliência emocional; O: Abertura à novas

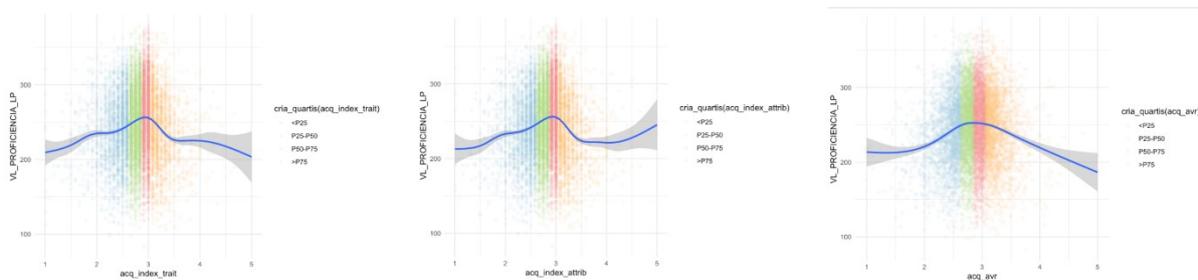
ideias; 'lw' – situação da Viheta-âncora de traço baixo; 'hi' – situação da Viheta-âncora de traço alto; '0' – situação de Viheta-âncora de polo negativo; '1' – situação de Viheta-âncora de polo positivo.

Os resultados observados na Tabela 11 demonstraram que o grau de associação entre o índice geral de aquiescência no atributo e no traço foi quase perfeito, com $r = 0,96$. Isso sugere que ambos os índices de aquiescência nas vinhetas são equivalentes não importando a maneira de se calcular via traço ou atributo.

Novamente, um conjunto de correlações de Pearson foi utilizado com o intuito de investigar as associações dos índices de aquiescência com variáveis externas, tais como sexo, série escolar, notas padronizadas de Linguagens e Matemática e o desempenho nas dimensões socioemocionais. Uma série de visualizações gráficas foram feitas no intuito de compreender estas relações. Assim, a Figura 6 traz algumas visualizações entre a relação dos índices de aquiescência com algumas das variáveis externas investigadas.

Figura 6

Visualização da relação entre nota de Linguagens com os índices de aquiescência



Nota. Na primeira imagem, temos a relação de Linguagens com o índice geral de aquiescência no traço; a segunda imagem traz a relação da nota de Linguagens com o índice geral de aquiescência no atributo; a terceira imagem, temos a relação da nota de Linguagens com o índice global de aquiescência.

A partir das imagens expostas na Figura 6, percebe-se que existe uma relação não-linear entre os índices de aquiescência com as notas em Linguagens, uma vez que existe uma concentração das notas de Linguagens (eixo X) na região do meio (isto é, valores próximos a

3, no eixo Y). Ou seja, o valor do índice próximo de três indicou respostas esperadas, ou consistentes. À medida que se afastaram, isso indicou a presença de respostas incoerentes, seja por aquiescência ou disaquiescência. Estes casos estão relacionados negativamente às notas.

Ao padronizar os índices, ou seja, subtrair a chave de valor três, conseguiu-se resultados mais fidedignos da relação entre eles com as notas padronizadas e demais variáveis externas. A Tabela 12 demonstra esses resultados.

Tabela 12

Correlações entre notas padronizadas de Linguagens e Matemática, gênero e séries escolares com os índices de aquiescência padronizados

	Linguagens	Matemática	Gênero	Série geral	EF_5	EF_9	EM_1
acq_index_trait	-.17**	-.16**	-.02	-.03	.04	-.04	.01
acq_index_attrib	-.18**	-.17**	-.03	-.03	.04	-.03	.00
acq_avr	-.17**	-.15**	-.02	-.11**	.11**	-.05*	-.03
acq_sd	-.03	-.01**	-.06**	.00	.01	-.04	.04
A_hi_Paula	-.21**	-.16**	-.07*	.00	.01	-.04	.04
A_lw_Julia	-.15**	-.13**	-.04	.01	.00	-.01	.02
C_hi_Juliana	-.21**	-.19**	-.07**	-.00	.01	-.01	.01
C_lw_Aline	-.12**	-.12**	-.01	-.01	.01	-.01	.01
E_hi_Paulinho	-.22**	-.18**	-.07**	-.07**	.08**	-.04	-.01
E_lw_Tiago	-.15**	-.11**	-.04*	-.03	.05*	-.07*	.04
N_hi_Pedro	-.21**	-.17**	-.02	-.01	.02	-.04	.02
N_lw_Beto	-.22**	-.18**	-.02	-.04*	.05*	-.05*	.01
O_hi_Amanda	-.18**	-.14**	-.02	-.06*	.06*	.00	-.04
O_lw_Luciana	-.12**	-.09**	-.03	-.05*	.04	.00	-.03
A_0_briguenta	-.16**	-.14**	-.04	-.10**	.10**	-.03	-.04
A_1_amável	-.21**	-.17**	-.03	-.06*	.06*	-.03	-.02
C_0_desleixada	-.13**	-.12**	-.03	.00	.02	-.05*	.04
C_1_organizada	-.17**	-.17**	-.04**	.02	-.03	.01	.01
E_0_tímido	-.20**	-.17**	-.07**	-.06*	.07**	-.05*	-.00
E_1_desinibido	-.19**	-.17**	-.03	-.11**	.12**	-.08**	-.01
N_0_nervoso	-.22**	-.17**	-.04	-.06*	.07**	-.05*	.00
N_1_calmo	-.22**	-.18**	-.04	-.07**	.08**	-.06*	.00
O_0_tradicional	-.09	-.09**	.01	-.04	.04	.00	-.03
O_1_criativa	-.16**	-.13**	-.02	-.05*	.06*	-.05*	.01

Nota, * $p < 0,05$; ** $p < 0,01$; Série geral: alunos do 5º e 9º ano do Fundamental I e 1º do Ensino Médio; EF_5 – alunos do 5º ano do Fundamental I; EF_9 – alunos do 9º ano do Fundamental I; EM_1 – 1º ano do Ensino Médio; acq_index_trait – índice de aquiescência no traço; acq_index_attrib – índice de aquiescência no atributo; acq_avr

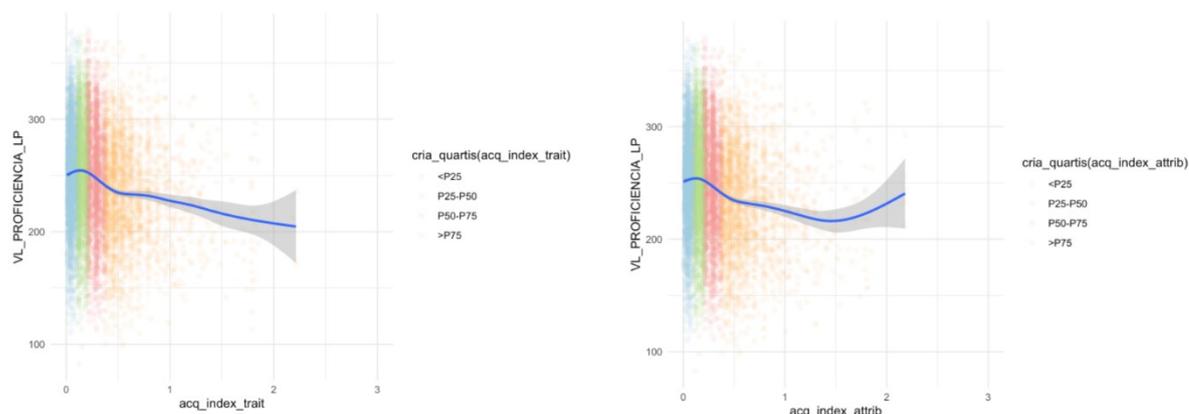
– média do índice de aquiescência; acq_sd – desvio-padrão do índice de aquiescência; A: Amabilidade; C: Autogestão; E: Engajamento com os outros; N: Resiliência emocional; O: Abertura à novas ideias; ‘lw’ – situação da Vinheta-âncora de traço baixo; ‘hi’ – situação da Vinheta-âncora de traço alto; ‘0’ – situação de Vinheta-âncora de polo negativo; ‘1’ – situação de Vinheta-âncora de polo positivo.

Os resultados apresentados na Tabela 12 demonstraram relações de grau fraco e direção negativa entre os índices de aquiescência e as notas padronizadas de Linguagens e Matemática. Contudo, não se tem este padrão com as variáveis sexo e séries escolares, mesmo com séries mais avançadas, como é o caso do 1º ano do ensino Médio (EM_1), que não apresentou nenhuma correlação significativa.

Com esses índices, novos gráficos de dispersão foram criados para visualização dos índices com as notas em Linguagens. A Figura 7 apresenta os gráficos do índice de aquiescência no traço e o índice de aquiescência no atributo criados.

Figura 7

Visualização da relação entre os índices de aquiescência no traço e no atributo com a nota em Linguagens



Nota. Na primeira figura, lado esquerdo, tem-se o gráfico de dispersão demonstrando a relação da nota de Linguagens e o índice de aquiescência padronizado no traço. Ao lado direito, o mesmo gráfico de dispersão com a nota de Linguagens com o índice de aquiescência padronizado no atributo.

Por fim, testou-se a associação dos índices de aquiescências com as dimensões socioemocionais avaliadas pelo SENNA. Aqui, também se utilizou os índices padronizados para a correlação de Pearson. A Tabela 13 demonstra os resultados.

Tabela 13

Correlação de Pearson entre os índices de aquiescência no traço, atributo, global e as dimensões socioemocionais

	C	E	N	A	O	NVLoc
acq_index_trait	-.02	-.07	-.02	-.07	-.07	.05
acq_index_attrib	-.02	-.08	-.02	-.08	-.08	.05
acq_avr	-.02	-.09	-.05	-.11	-.09	.05
acq_sd	.04	-.02	.07	-.03	.02	.02
A_hi_Paula	-.08	-.10	-.08	-.16	-.04	.13
A_lw_Julia	-.01	-.03	-.04	-.04	.01	.07
C_hi_Juliana	-.09	-.10	-.04	-.14	-.08	.10
C_lw_Aline	-.01	-.01	-.01	-.01	-.02	.03
E_hi_Paulinho	-.05	-.08	-.04	-.11	-.08	.09
E_lw_Tiago	-.02	-.06	-.03	-.08	-.06	.05
N_hi_Pedro	-.06	-.06	-.05	-.07	-.07	.09
N_lw_Beto	-.03	-.06	-.03	-.09	-.07	.10
O_hi_Amanda	-.03	-.07	-.03	-.09	-.06	.05
O_lw_Luciana	.00	-.01	-.03	.00	.02	.06
A_0_briguenta	.03	-.04	-.04	-.02	-.01	.09
A_1_amável	-.04	-.08	-.07	-.17	-.04	.14
C_0_desleixada	-.01	-.02	-.02	-.03	-.03	.04
C_1_organizada	-.11	-.09	-.05	-.10	-.08	.08
E_0_tímido	-.05	-.09	-.04	-.11	-.07	.09
E_1_desinibido	-.02	-.08	-.04	-.08	-.06	.09
N_0_nervoso	-.03	-.08	-.04	-.09	-.07	.09
N_1_calmo	-.03	-.08	-.02	-.10	-.07	.09
O_0_tradicional	.01	-.02	-.02	.01	.01	.04
O_1_criativa	-.02	-.04	-.03	-.05	-.06	.04

Nota. C – Autogestão; E – Engajamento com os outros; N – Resiliência emocional; A – Amabilidade; O – Abertura à novas ideias; NVLoc – Locus de controle negativo; acq_index_trait – índice de aquiescência no traço; acq_index_attrib – índice de aquiescência no atributo; acq_avr – média do índice de aquiescência; acq_sd – desvio-padrão do índice de aquiescência; ‘lw’ – situação da Vinheta-âncora de traço baixo; ‘hi’ – situação da Vinheta-âncora de traço alto; ‘0’ – situação de Vinheta-âncora de polo negativo; ‘1’ – situação de Vinheta-âncora de polo positivo

Os resultados demonstrados na Tabela 13 apontaram que os índices de aquiescência no atributo, traço e global têm maior associação com as dimensões Abertura à novas ideias, Amabilidade e Engajamento com os outros, com direção negativa. Um outro resultado que também chama a atenção é que os graus de associação com as dimensões socioemocionais foram vistos com os índices de aquiescência nas situações em que o polo é positivo. Por sua vez, as situações com polo negativo apresentaram graus de associação de graus menores, e alguns de forma não significativa. Por fim, as relações observadas com a dimensão locus de controle apresentaram graus fracos e direção positiva.

Mensuração de viés de grupo (*Group reference bias*; GRB) por Vinhetas-âncora

Por fim, as últimas análises referem-se à investigação de evidências de validade do uso de Vinhetas-âncora (VAs) para mensurar o erro do tipo sistemático conhecido como viés de grupo (GRB; *group reference bias*, em inglês). Para isso, foi calculado a média e o desvio-padrão de cada conjunto das VAs para visualizar o quanto a média está centrada em três, ou se há uma assimetria na distribuição das médias – ou seja, um deslocamento para direita ou esquerda – da mesma forma que se buscou compreender a dispersão do desvio-padrão do valor esperado de um.

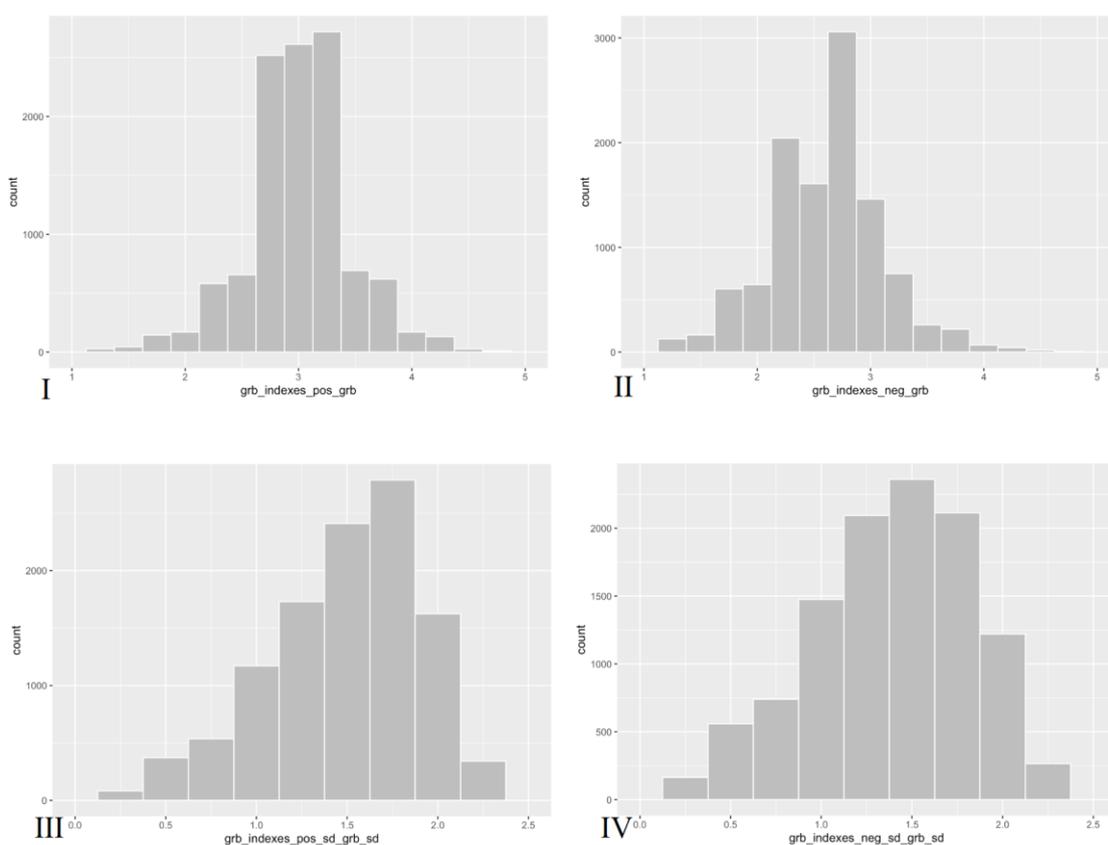
Estes índices foram calculados de forma geral, ou seja, há um índice que calcula a média (*grb_index_pos*) de todos os conjuntos de VAs em que se pediu para avaliar os personagens das histórias com um adjetivo de polo positivo, bem como o desvio-padrão dessas respostas (*grb_index_pos_sd*). Estes mesmos índices foram criados para as situações em que se pediu para avaliar os personagens das histórias com um adjetivo de polo negativo (*grb_index_neg* e *grb_index_neg_sd*, respectivamente). Contudo, neste último caso, as respostas foram invertidas. Isso permitiu chegar a um escore mais alto para as Vinhetas-âncora negativas pela negação ao atributo avaliado, permitindo uma comparação mais clara entre as médias e desvios-

padrão dos índices gerais de viés de grupo. Ainda, as médias e desvios-padrão para cada dimensão foram calculados.

Com estes índices calculados, o primeiro resultado a se mostrar é a distribuição dos quatro índices gerais, sendo eles o índice a partir da média geral das VAs de polo positivo, o índice com a média geral das VAs de polo negativo, o índice de desvio-padrão das VAs de polo positivo e, por fim, o índice de desvio-padrão das VAs de polo negativo. A Figura 8 mostra os histogramas desses índices.

Figura 8

Histogramas com as frequências de distribuição de cada índice geral calculado



Nota. O histograma I demonstra a frequência de distribuição do índice geral com as médias de viés de grupo para as Vinhetas-âncora de polo positivo. O histograma II, a frequência de distribuição do índice geral com as médias de viés de grupo para as Vinhetas-âncora de polo negativo. O histograma III apresenta a frequência para o índice de desvio-padrão de Vinhetas-âncora de polo positivo e o histograma IV, de polo negativo.

Os histogramas apresentados na Figura 8 trouxeram informações importantes sobre os índices. É possível perceber que o histograma I, relativo ao índice da média geral para viés de grupo das VAs de polo positivo, apresentou um formato simétrico e em torno do valor três. Contudo, no histograma II, referente ao índice da média geral para o conjunto de VAs de polo negativo, percebe-se que ele está puxado à esquerda, tendo uma média menor. Os histogramas III e IV, referentes aos índices de viés de grupo a partir dos desvios-padrão dos conjuntos de polo positivo e negativo, respectivamente, apresentaram um formato simétrico. Entretanto, os valores ficaram mais concentrados acima do valor de um, mais especificamente entre 1,25 e 1,75.

Em seguida, uma correlação de Pearson foi utilizada para averiguar a direção e o grau de associação dos índices de viés de grupo com as notas de Linguagens e Matemática. A Tabela 14 apresenta o resultado das correlações entre todos os índices calculados, ou seja, os gerais e os específicos de cada dimensão e as variáveis notas.

Tabela 14

Correlação entre os índices gerais e por dimensão de viés de grupo com notas em Linguagens e Matemática.

	Linguagens	Matemática
grb_indexes_pos_grb	.07*	.07*
grb_indexes_neg_grb	-.02	.01
grb_indexes_pos_sd_grb_sd	.23*	.20*
grb_indexes_neg_sd_grb_sd	.25*	.21*
A_0_grb	-.02	-.03
A_1_grb	.02	.07*
C_0_grb	-.01	-.02
C_1_grb	.06*	.05*
E_0_grb	.03	.01
E_1_grb	.13*	.12*
N_0_grb	.03	.01
N_1_grb	.11*	.10*
O_0_grb	-.04	-.05*
O_1_grb	-.06*	-.03
A_0_grb_sd	.19*	.14*
A_1_grb_sd	.14*	.09*
C_0_grb_sd	.16*	.15*
C_1_grb_sd	.16*	.15*
E_0_grb_sd	.20*	.17*
E_1_grb_sd	.18*	.16*
N_0_grb_sd	.23*	.17*
N_1_grb_sd	.22*	.18*
O_0_grb_sd	.16*	.12*
O_1_grb_sd	.13*	.11*

Nota: * $p < 0,05$; 'grb' – Média da referência de grupo; 'grb_sd' – Desvio-padrão da referência de grupo; 'pos' – positivo; 'neg' – negativo; A – Amabilidade; 'C' – Autogestão; 'E' – Engajamento com os Outros; 'N' – Resiliência Emocional; 'O' – Abertura à Novas Ideias.

O índice geral da média de viés de grupo do conjunto de polo positivo apresentou correlações significativas e de direção positiva com as notas de linguagem e Matemática, o que não pôde ser observado para o índice da média do conjunto de polo negativo. Contudo, ambos os índices de desvio-padrão de viés de grupo apresentaram resultados significativos e de direção positiva, com grau alto de relação. Olhando os índices da média por dimensão de viés de grupo, percebe-se que os índices da média do conjunto de VAs de polo negativo para as dimensões Autogestão, Engajamento com os outros e Resiliência emocional não tiveram resultados

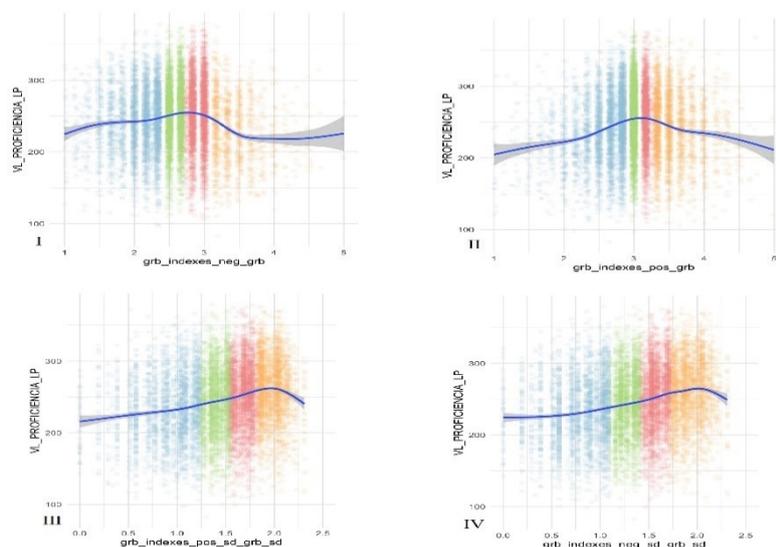
significativos. Por sua vez, os índices da média do conjunto de polo positivo para estas mesmas dimensões apresentaram resultados significativos e de direção positiva para ambas as notas de Linguagens e Matemática, em especial a dimensão Engajamento com os Outros, que apresentou o maior grau de relação com estas variáveis.

Ainda sobre os resultados apresentados na Tabela 14, os índices da média de viés de grupo do conjunto negativo de Amabilidade e Abertura à Novas Ideias apresentaram uma relação significativa com a nota de Matemática, mas não tiveram resultados significativos com as notas de Linguagens. O índice da média de viés de grupo de polo positivo de Abertura à Novas Ideias, por sua vez, apresentou uma relação significativa com a nota de Linguagens, mas não apresentou uma relação significativa com a nota de Matemática. Por fim, todos os índices de desvios-padrão de viés de grupo apresentaram relações significativas e direções positivas, ressaltando para as dimensões Engajamento com os outros e Resiliência emocional, que tiveram os maiores graus de relação.

De maneira similar aos Índices de Aquiescências, visualizações gráficas foram criadas. Isto tem como objetivo compreender mais detalhadamente como as relações acontecem e se lineares ou não, o que pode explicar os graus de relações apresentados pela correlação de Pearson, bem como os graus desta associação. A Figura 9 apresenta essas visualizações.

Figura 9

Visualizações gráficas da dispersão entre as notas de Linguagens e os índices gerais das médias e desvios-padrão de viés de grupo



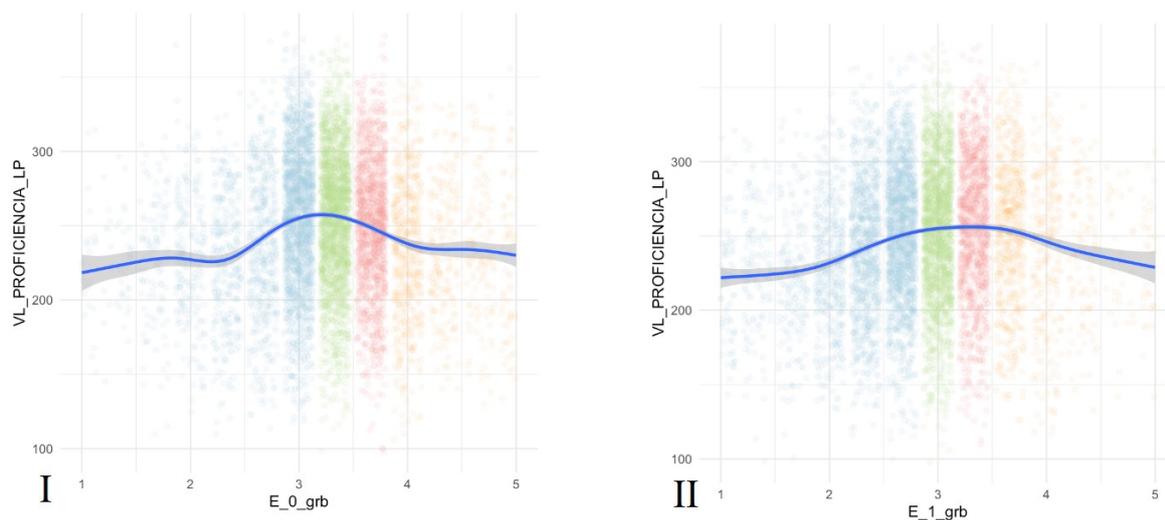
Nota. O gráfico I mostra a dispersão entre as notas em Linguagens e o índice de média geral de polo positivo de viés de grupo. O gráfico II mostra a dispersão entre a nota em Linguagens e o índice de média geral de polo negativo de viés de grupo. O gráfico III mostra a dispersão entre a nota em Linguagens e o índice de desvio-padrão de polo positivo de viés de grupo. O gráfico IV mostra a dispersão entre a nota em Linguagens e o índice de desvio-padrão de polo negativo de viés de grupo.

A partir dos gráficos na Figura 9, faz-se algumas observações destes resultados. O índice de média de viés de grupo de polo positivo apresentou uma relação não-linear, uma vez que existiu uma concentração dos valores deste índice no eixo X na região mais ao meio (valores próximos a três). Para os índices de desvios-padrão, ambos apresentaram uma dispersão crescente entre as notas em Linguagens e os valores obtidos nestes índices.

Os resultados na Tabela 14 mostraram que a dimensão Engajamento com os Outros apresentou uma correlação significativa entre esta dimensão e as notas para o índice de média do polo positivo, mas não para o índice de média de polo negativo. Desta forma, recorreu-se a visualizações gráficas para compreender melhor a relação da dispersão existente entre essas duas variáveis. A Figura 10 demonstra essa relação.

Figura 10

Visualização gráfica da dispersão entre as notas em Linguagens com os índices de média de viés de grupo para a dimensão Engajamento com os outros



Nota. No gráfico I, apresenta-se a dispersão entre notas em Linguagens e o índice de média de polo negativo de viés de grupo para a dimensão Engajamento com os Outros. No gráfico II, apresenta-se a dispersão entre notas em Linguagens e o índice de média de polo positivo de viés de grupo para a dimensão Engajamento com os Outros.

É possível observar na Figura 10 que a relação entre as notas em Linguagens e ambos os índices de médias de viés de grupo para a dimensão Engajamento com os Outros não foi linear. No primeiro gráfico foi possível observar que houve uma queda no grau da associação, seguido de repentino aumento na associação entre o índice e as notas em Linguagens. Contudo, essa relação voltar a declinar. No segundo gráfico, houve uma tendência a uma associação positiva, com uma queda para valores mais altos do índice.

Continuando a investigar a relação dos índices de viés de grupo com variáveis externas, uma correlação de Pearson foi utilizada para averiguar a direção e o grau de associação estes índices com as variáveis sociodemográficas, a saber, sexo, série escolar geral e as séries escolares de forma específica. A Tabela 15 apresenta o resultado das correlações entre todos os índices calculados, ou seja, os gerais e os específicos de cada dimensão e as variáveis sociodemográficas.

Tabela 15

Correlação entre os índices gerais e por dimensão de viés de grupo com gênero, série escolar geral e séries escolares específicas.

	Gênero	Série geral	EF_5	EF_9	EM_1
grb_indexes_pos_grb	-.01	.03	-.04	.03	-.00
grb_indexes_neg_grb	-.04	-.04	.03	.02	-.04
grb_indexes_pos_sd_grb_sd	.08*	-.01	-.00	.04	-.04
grb_indexes_neg_sd_grb_sd	.06*	.00	-.02	.05*	-.04
A_0_grb	.03	-.07*	.07*	-.02	-.03
A_1_grb	.04	-.06*	.07*	-.05*	-.01
C_0_grb	.03	.03	-.02	-.01	.03
C_1_grb	-.01	.01	-.02	.03	-.02
E_0_grb	.03	.04	-.04	.01	.02
E_1_grb	-.05*	.10*	-.11*	.06*	.02
N_0_grb	.01	.03	-.02	-.02	.03
N_1_grb	-.01	.04	-.04	.02	.01
O_0_grb	.04	-.05*	.06*	-.05*	.01
O_1_grb	-.00	-.13*	.14*	-.06*	-.04
A_0_grb_sd	.02	.07*	-.07*	.03	.02
A_1_grb_sd	.06*	-.04	.04	.00	-.03
C_0_grb_sd	.04	-.06*	.05*	.03	-.06*
C_1_grb_sd	.07*	-.05*	.05*	.01	-.04
E_0_grb_sd	.08*	.04	-.05*	.05*	-.02
E_1_grb_sd	.05*	.03	-.04	.05*	-.02
N_0_grb_sd	.05*	.03	-.04	.06*	-.03
N_1_grb_sd	.07*	.03	-.05*	.07*	-.03
O_0_grb_sd	.02	-.00	-.00	.03	-.03
O_1_grb_sd	.04	-.01	-.00	.04	-.03

Nota: * $p < 0,05$;

Como é possível visualizar a Tabela 15, os índices gerais de média de viés de grupo não apresentaram correlações estatisticamente significativas com as variáveis sociodemográficas. Já os índices gerais de desvio-padrão de viés de grupo apresentaram correlações estatisticamente significativas com as variáveis gênero. Chama a atenção que o índice geral de desvio-padrão de polo negativo apresentou uma correlação estatisticamente significativa com a série escolar referente ao 9º ano do Ensino Fundamental I. Os índices da média por dimensão que apresentaram alguns resultados estatisticamente significativos foram Amabilidade e

Abertura à novas ideias, ambos os polos, e Engajamento com os Outros, apenas para o polo negativo. Já os índices de desvio-padrão por dimensão, a dimensão Abertura à Novas Ideias foi a única a não apresentar resultados estatisticamente significativo para nenhum dos polos.

A última análise realizada foi uma correlação de Pearson dos índices de viés de grupo com os fatores socioemocionais avaliados pelo SENNA. A Tabela 16 apresenta os resultados encontrados.

Tabela 16

Correlação entre os índices gerais e por dimensão de viés de grupo com as dimensões socioemocionais avaliados pelo SENNA.

	C	E	N	A	O	NVLoc
grb_indexes_pos_grb	.06*	.12*	.03	.13*	.14*	.03
grb_indexes_neg_grb	-.01	.02	-.04	.03	.07*	.10*
grb_indexes_pos_sd_grb_sd	.12*	.18*	.08*	.22*	.16*	-.11*
grb_indexes_neg_sd_grb_sd	.08*	.17*	.06*	.18*	.16*	-.09*
A_0_grb	.07*	-.04	.04	.03	-.06*	-.08*
A_1_grb	.08*	.08*	-.04	.12*	.12*	.04
C_0_grb	.01	-.00	.02	-.02	-.05*	-.04
C_1_grb	.02	.07*	.01	.07*	.07*	.02
E_0_grb	-.03	-.01	.03	-.04	-.05*	-.08*
E_1_grb	-.02	.10*	.02	.07*	.11*	-.01
N_0_grb	.01	-.02	.08*	-.03	-.07*	-.10*
N_1_grb	.07*	.08*	.07*	.11*	.13*	.01
O_0_grb	.03	-.01	.04	-.02	-.09*	-.07*
O_1_grb	.12*	.07*	.03	.11*	.12*	.07*
A_0_grb_sd	.02	.13*	.06*	.13*	.11*	-.10*
A_1_grb_sd	.10*	.16*	.11*	.26*	.14*	-.12*
C_0_grb_sd	.08*	.10*	.04	.12*	.11*	-.07*
C_1_grb_sd	.13*	.12*	.06*	.14*	.11*	-.08*
E_0_grb_sd	.07*	.17*	.05*	.17*	.12*	-.08*
E_1_grb_sd	.06*	.16*	.05*	.16*	.13*	-.09*
N_0_grb_sd	.06*	.15*	.05*	.16*	.14*	-.07*
N_1_grb_sd	.09*	.15*	.07*	.18*	.13*	-.11*
O_0_grb_sd	.04	.09*	.01	.10*	.14*	-.05*
O_1_grb_sd	.06*	.09*	.04	.13*	.12*	-.05*

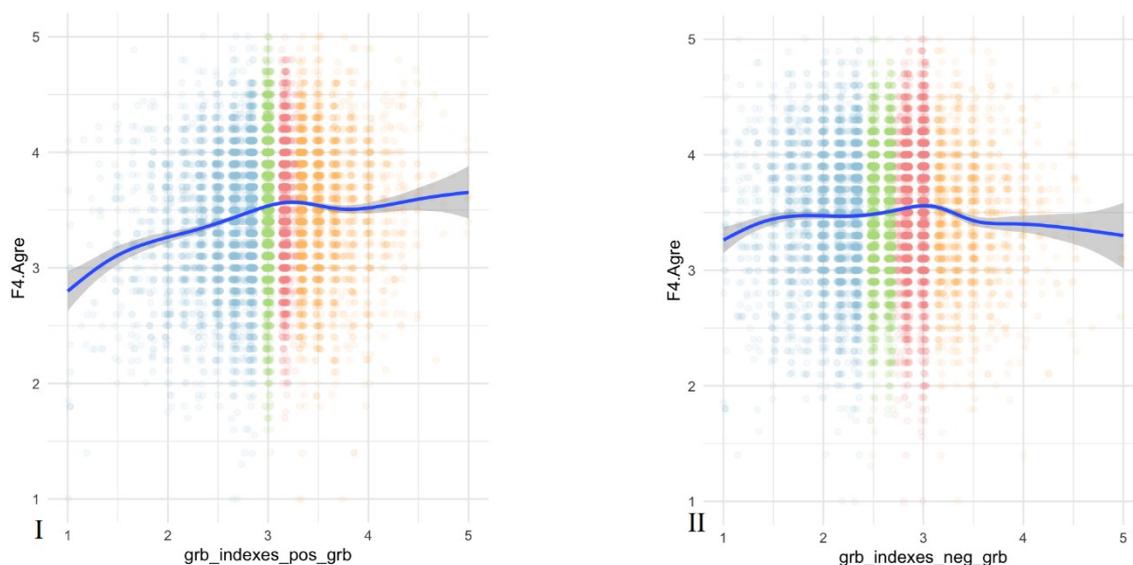
Nota. * $p < 0,05$; C – Autogestão; E – Engajamento com os outros; N – Resiliência emocional; A – Amabilidade; O – Abertura à novas ideias; NVloc – Lócus de controle negativo;

Os resultados apresentados na Tabela 16 demonstram que as dimensões socioemocionais se encontram, em sua maioria, com algum grau de relação, mas sem um padrão de direção. Dentre os resultados observados, destaca-se as dimensões Amabilidade e Abertura à Novas Ideias, que apresentaram um maior número de associações estatisticamente significativas com os diferentes índices de viés de grupo e maiores graus de associação. No tocante aos índices de média de viés de grupo, as dimensões que mais apresentaram relações estatisticamente significativas foram de Resiliência Emocional e Abertura à Novas Ideias, ambas de polo negativo.

Continuando com as visualizações gráficas, aqui se apresenta dois gráficos relativos à dispersão entre os índices gerais da média de viés de grupo de polo positivo e negativo com a dimensão Amabilidade. A Figura 11 apresenta esses gráficos.

Figura 11.

Visualização gráfica da dispersão entre os escores da dimensão Amabilidade com os índices gerais de média de viés de grupo



Nota. No gráfico I, apresenta-se a dispersão os escores da dimensão Amabilidade e o índice geral de média de polo positivo de viés de grupo. No gráfico II, apresenta-se a dispersão os escores da dimensão Amabilidade e o índice geral de média de polo negativo de viés de grupo.

A partir dos gráficos da Figura 11, foi possível perceber que o índice geral de média de viés de grupo de polo positivo tiveram uma concentração no valor médio de três, conforme esperado. Já o índice para o polo negativo se apresentou mais esparso e de forma não linear.

Os resultados aqui apresentados trouxeram diferentes tipos de informação. Na próxima seção, discutir-se-á seus possíveis significados. Assim, inicia-se relacionando cada resultado exposto com uma explanação plausível para seu acontecimento. Após todos os tópicos serem apresentados, explicações mais concisas serão expostas, a fim de consolidar as informações apresentadas.

Discussão

A discussão seguirá com a sequência em que foram apresentados, seguidos de uma visão geral acerca do uso de Vinhetas-âncora com mensuração de erros do tipo sistemático, em especial, os estilos de resposta. Assim, inicia-se a discussão com o primeiro tópico sobre as análises descritivas do uso da escala de resposta das VAs. Depois, tem-se a relação das VAs com fatores externos, a saber, desempenho em um instrumento de autorrelato que avalia as competências socioemocionais, o SENNA, variáveis sociodemográficas (gênero e série escolar), além das variáveis critérios que são notas do desempenho em Linguagens e Matemática. Nestes pontos, vale ressaltar de que como se trata de um estudo inédito, a discussão dos resultados é feita com base no entendimento que a literatura da área traz sobre como estilos de resposta impactam o endosso aos itens, bem como na interpretação aqui gerida pelo autor da tese e orientador.

Dando continuidade, o terceiro tópico discute os resultados do uso de Vinhetas-âncora para mensurar o estilo de resposta aquiescente, enquanto o quarto tópico traz um debate acerca do uso das VAs para mensuração do estilo de resposta de viés de grupo. Por fim, uma visão geral sobre os resultados discutidos é apresentada.

Uso da escala de resposta das Vinhetas-âncora

O objetivo deste estudo foi verificar como o uso das Vinhetas-âncora pode ser útil para identificar erros sistemáticos causados por estilos de respostas. Desta forma, este estudo conseguiu demonstrar evidências de validade iniciais sobre o uso de VAs como uma possível técnica para controle destes vieses. Para isto, foi necessário que um conjunto de vinhetas de polo negativo fosse testado, pela primeira vez, e compará-la com o uso de vinhetas de polo positivo. Isto permitiu analisar se as mesmas pessoas respondem de maneira semelhante quando o mesmo estímulo é apresentado de maneira inversa (i.e., em um polo distinto do que se é

comumente utilizado), de forma que uma inconsistência neste padrão de resposta pode indicar uma tendência ao uso de estilos de resposta. Especificamente, buscou-se avaliar que tipos de estilos de respostas podem ser medidos e/ou controlados com o uso desta técnica, salientando que poucos estudos olharam para as VAs com esta finalidade (cf Kyllonen & Bertling, 2014), colocando o erro sistemático apenas como DIF (He et al., 2017; Weiss & Roberts, 2018).

Em suma, com os resultados apresentados, podemos inferir que, por um lado, quando uma situação é apresentada no polo positivo, os respondentes tendem a endossar com maior frequência e de maneira esperada, havendo maior coerência nas respostas. Por outro lado, as situações apresentadas no polo negativo tendem a ter respostas menos coerentes, impactando diretamente na frequência observada das respostas dadas. Assim, percebe-se que as Vinhetas-âncora de polo negativo apresentam uma maior variabilidade na frequência de respostas dadas pelos respondentes.

Estes resultados podem indicar uma interação entre a polaridade das Vinhetas-âncora com os adjetivos utilizados para avaliar os domínios. Isto é, existe uma maior frequência em discordar quando a pessoa apresenta um alto nível em um determinado traço e se questiona o quanto baixo aquele traço está presente (por exemplo, avaliar o quanto uma pessoa organizada é bagunceira), do que em concordar quando a pessoa apresenta um nível baixo daquele traço no mesmo contexto (e.g., avaliar o quanto uma pessoa bagunceira é bagunceira). Neste sentido, o estudo de Primi et al. (2016b) achou resultados semelhantes, em que os escores ajustados por Vinhetas-âncora trouxeram uma melhora na consistência interna do instrumento ao reduzir o erro sistemático, mas como o conjunto de situações de polo negativa não havia sido testada, maiores inferências não foram realizadas, o que se pode perceber no resultado aqui apresentado.

Ou seja, adjetivos negativos eliciam uma “discordância geral” maior mesmo quando apropriados (bagunceira com uma pessoa que seja bagunceira) e ainda mais forte em situações dissonantes aparentemente injustas de atribuição de algo negativo a uma pessoa com

características positivas. Alguns estudos com o autorrelato demonstraram que algo nesse sentido pode acontecer, em especial quando se tem instrumentos com pares semânticos (Soto & John, 2017; Primi et al., 2021; Wetzell et al., 2016).

Relação das Vinhetas-âncora com fatores externos

A partir das correlações apresentadas entre as situações das VAs e as notas entre Linguagens e Matemática, é possível perceber que existe uma associação entre o desenvolvimento cognitivo e o uso das Vinhetas-âncora. Em geral as respostas coerentes, isto é, endosso de característica de polo positivo para personagens com traço positivo (ou vice-versa, endosso de características negativas para pessoas com traços negativos) tem correlação positiva com desempenho. Por outro lado, respostas incoerentes, ou seja, o endosso de característica de polo positivo para personagens com traço negativo (ou o contrário, endosso de características negativas para pessoas com traços positivos) tem correlação negativa com desempenho.

Nota-se que as vinhetas que descrevem o mesmo comportamento, mas que estão em polos opostos, tendem a apresentar um grau menor de correlação em detrimento às vinhetas de polo positivo. Uma explicação plausível para isso é que uma melhor congruência para o endosso às Vinhetas-âncora está relacionada diretamente com a capacidade cognitiva da pessoa em compreender a tarefa. Em outras palavras, quanto maior o nível de capacidade cognitiva (isto é, inteligência), maior o nível de congruência para as respostas das VAs (He et al., 2017; Lechner & Rammstedt, 2015; McLarnon & Carswell, 2013; Vigil-Colet et al., 2012; Wetzell et al., 2016; Zanon et al., 2018).

Ainda, percebe-se que o sexo pode ser um fator que também pode influenciar respostas mais congruentes. Ao se olhar os resultados de correlação das Vinhetas-âncora com esta variável, o resultado de direção negativa demonstra que o gênero feminino apresenta essa

capacidade mais desenvolvida. Ou seja, elas podem ter uma maior capacidade verbal desenvolvida, o que facilitaria em compreender a tarefa, e, com isso, fazer um melhor uso da escala de resposta tanto das situações das Vinhetas-âncora quanto de um teste de autorrelato (Lelkes & Weiss, 2015; Soto et al., 2008, 2011).

Uma questão importante é o quanto a correlação das competências socioemocionais com desempenho pode estar confundida pela habilidade cognitiva, já que responder a um teste depende da habilidade cognitiva. Como demonstrado nas correlações com as dimensões do SENNA, competências socioemocionais mais bem desenvolvidas tem uma relação positiva com melhor uso das escalas de respostas. Esse debate, com resultados semelhantes, pode ser visto em diferentes estudos ao longo do tempo (Crobach, 1942; Marsh, 1987; Mõttus et al., 2012; Paulhus, 1991; Wetzel & Carstensen, 2017).

Uso de Vinhetas-âncora para mensurar aquiescência

Traço de aquiescência e variância no traço apresentam correlação forte, então é possível inferir que variações nos traços avaliados estão mais relacionados com um endosse mais aquiescente – semelhante à variância no atributo. Isto pode ser um indicativo de resposta aquiescente, já que itens negativos requerem uma habilidade cognitiva mais refinada para um endosso mais consistente em relação ao item e chave de resposta (Rammstedt & Farmer, 2013; Wetzel et al., 2016; Zanon et al., 2018). Isto é corroborado pela literatura, ao conseguir demonstrar evidências que responder a itens negativos requerem um esforço cognitivo maior (He et al., 2017; Soto et al., 2011; Wetzel et al., 2016).

Contudo, o mesmo não pôde ser observado nos itens de vinhetas âncoras de polo positivo. Em suma, os itens destes conjuntos tiveram uma distribuição dentro do esperado para cada situação, indicando que concordar com os diferentes níveis de traços latentes não apresenta diferentes graus de endosso quando se pergunta de maneira positiva. Afinal, sabe-se que itens

de polo positivo tendem a ser de entendimento mais simples mesmo para pessoas com menor habilidade cognitiva (Cronbach, 1946; Wetzel et al., 2016).

Os diferentes índices de aquiescência se apresentaram significativos. Em suma, estão relacionados com uma capacidade cognitiva. É possível inferir que algumas dimensões socioemocionais podem estar mais relacionadas à uma tendência de resposta aquiescente, em especial Amabilidade e Abertura à Novas Ideias. Importante ressaltar que a relação entre o índice de consistência do conjunto positivo apresentou os mesmos valores de r e mesma direção para Linguagens e Matemática, enquanto o índice geral apresentou valores de r também iguais para estas duas medidas, com uma direção positiva e outra negativa. Estima-se que este resultado se dê pelo fato que o índice geral seja calculado pelas médias do índice de conjunto de vinhetas positivas e negativas, sendo que este último conjunto não apresentou resultados significativos para ambas as medidas de desempenho. Isto pode ser explicado pela dificuldade em responder itens negativos por crianças e adolescentes (Primi et al., 2016a). Outra possibilidade é a não linearidade (distribuição assimétrica) que os índices apresentam, de forma que não se torna possível observar o real desempenho dos alunos com índices de consistências mais próximos ao valor 3.

Uso de Vinhetas-âncora para mensurar viés de grupo

Para investigar se as Vinhetas-âncora poderiam ser utilizadas como forma de mensuração para o erro sistemático do estilo de resposta de viés de grupo, ambos os conjuntos de polos foram utilizados. Ressalta-se que outros estudos já se utilizaram das VAs para tentar mensurar e corrigir o viés de grupo, alguns com resultados satisfatórios (Kyllonen & Bertling, 2014; Primi et al., 2016b; Vonkova et al., 2015; Xu & Xie, 2016), outros com resultados contrários ao uso da técnica (Bago d'Uva et al., 2008; He et al., 2017). Contudo, o uso de um conjunto de situações de polo negativo foi, pela primeira vez, testado aqui.

Os resultados descritivos demonstraram que os índices de viés de grupo criado pelo uso dos conjuntos de VAs se apresentaram de forma simétrica, ou seja, demonstrando que há um padrão de uso dentro do que se espera. Tal resultado pode ser explicado pelo que já foi discutido aqui, que o uso de uma escala de resposta está associado com um maior desenvolvimento acadêmica e cognitivo. Assim, não se parece haver um uso enviesado a partir das diferenças de grupo, pela dispersão vista.

Ao observar os resultados das correlações dos índices de referência de grupo com o desempenho escolar, a partir das notas de Linguagens e Matemática, os índices calculados com os desvios-padrão se apresentaram estatisticamente significativos. Isso pode ser uma evidência de que quanto maior a dispersão no uso das escalas de resposta, uma maior habilidade cognitiva pode ser encontrada. Assim, com base nas pesquisas de He e van de Vijer (2016) e Yousfi (2018), os resultados do nosso estudo corroboram estes achados de que existe uma dimensão cognitiva para o uso de referência ao grupo. Em outras palavras, é possível afirmar que os resultados dos índices das médias apresentaram um padrão similar aos índices de aquiescência. Soma-se a isso que os resultados dos desvios-padrão têm relação positiva. Tais resultados estão indicando que existem padrões consistentes de resposta (i.e., no uso da escala), de forma que estes índices conseguem mensurar a variância no atributo, que se correlaciona com desempenho escolar, reforçando a característica cognitiva do uso de Vinhetas-âncora.

Crane et al. (2015), Léon et al. (2014) e Weiss e Roberts (2018) apresentaram resultados de que o uso de Vinhetas-âncora para controlar erros do tipo sistemático em variáveis sociodemográficas pode ser uma técnica promissora. Ao testarmos essa hipótese, percebemos que os índices de média para o viés de grupo são quase nulos, mas os índices de desvio-padrão são estatisticamente significativos. Em especial, as meninas apresentaram uma maior dispersão no uso de Vinhetas-âncora, enquanto alunos em séries menores apresentaram uma dispersão menor. Esses resultados indicam que as meninas podem apresentar respostas mais consistentes,

ao se levar com conta o sexo, e que quanto menor a série, maior uma tendência de se responder de acordo com um grupo de referência. Isto pode ser devido ao fato de que o processo maturacional e a necessidade de se estar em um grupo influencia pessoas em uma série escolar menor (Soto et al., 2011).

Visão geral acerca do uso de Vinhetas-âncora para mensurar erros do tipo sistemático

O primeiro resultado obtido demonstra que as VAs do conjunto de polo negativo apresentam uma tendência de serem endossados com a extremidade negativa. Isto pode ser um indicativo de resposta aquiescente, já que itens negativos requerem uma habilidade cognitiva mais refinada para um endosso mais consistente em relação ao item e chave de resposta (Rammstedt & Farmer, 2013; Wetzel et al., 2016; Zanon et al., 2018). Isto é corroborado pela literatura, ao conseguir demonstrar evidências que responder a itens negativos requerem um esforço cognitivo maior (He et al., 2017; Soto et al., 2011; Wetzel et al., 2016).

Contudo, o mesmo não pôde ser observado nos itens de Vinhetas-âncoras de polo positivo. Em suma, os itens destes conjuntos tiveram uma distribuição dentro do esperado para cada situação, indicando que concordar com os diferentes níveis de traços latentes não apresenta diferentes graus de endosso quando se pergunta de maneira positiva. Afinal, sabe-se que itens de polo positivo tendem a ser de entendimento mais simples mesmo para pessoas com menor habilidade cognitiva (Cronbach, 1946; Wetzel et al., 2016).

De maneira semelhante, os resultados dos histogramas para as médias por domínio indicaram que o conjunto de vinhetas de polo positivo teve curvas de distribuição como esperado. Vale ressaltar que a situação do nível de traço médio apresentou uma curva simétrica, mas mesocúrtica ao invés de leptocúrtica. Isto pode ser explicado pelo fato de que as chaves centrais tendem a ser alvo de dúvidas para os respondentes, em especial na faixa da

adolescência, de forma que se torna aceitável que esta distribuição não comprometa o resultado achado (Savalei & Falk, 2014; Soto et al., 2011).

Ainda, as análises de correlação das situações das Vinhetas-âncora, junto com os demais índices criados para este estudo (índices de consistência, aquiescência, viés de grupo), demonstraram que o erro sistemático está presente em associações com diferentes fatores externos. Especialmente, a capacidade cognitiva parece estar fortemente relacionada com respostas mais aquiescentes, ou disaquiescentes, já que compreender o uso da escala requer um esforço cognitivo (Soto et al., 2008, 2011; Primi et al., 2016a, 2016b, 2019, 2021; Wetzel et al., 2016; Zanon et al., 2018). Também foi possível observar que a variância no atributo está relacionada com habilidades cognitivas, associando o viés de grupo com este fator. Por último, possuir um repertório verbal está positivamente associado com essa capacidade cognitiva, que permite compreender melhor a tarefa do instrumento de autorrelato, gerando uma maior consistência na resposta. Desta forma, ter competências socioemocionais mais bem desenvolvidas, em especial as dimensões Abertura à Novas Ideias e Amabilidade, encontram resultados mais constantes com o uso mais coerente da escala de resposta (Kyllonen & Bertling, 2014; McCrae, 2018; Primi et al., 2021; Tracey, 2016).

Considerações finais

Diante do exposto, este estudo conseguiu demonstrar que o uso das âncoras vinhetas pode ser uma técnica útil para a mensuração de estilos de respostas como fonte do erro de tipo sistemático. Em especial, controle para os estilos de respostas aquiescente pode ser diretamente beneficiado com o uso de Vinhetas-âncora em conjunto de itens de polo positivo e negativo. Contudo, esta pesquisa não está isenta de limitações. Inicialmente, avaliações mais completas, envolvendo fatores de personalidade mais tradicionais e medidas externa de desempenho cognitivo (inteligência, memória, raciocínio etc.) não foram utilizados, limitando as inferências acerca da validade critério aqui encontradas. Ademais, apesar do número amostral grande, este estudo foi limitado a apenas um estado, que pode não replicar em demais localidades, bem como se os mesmos resultados seriam aplicados para população adulta, idosa, grupos clínicos, contextos compulsórios, dentre outros.

Sugere-se que estudos futuros possam avaliar como o uso das Vinhetas-âncora se utilizem de instrumentos de medidas conhecidamente impactadas pelos estilos de respostas, a fim de que se possa comparar suas propriedades psicométricas antes e após a correção, bem como buscar replicar os índices aqui propostos. O primeiro ponto é a indicação de que a técnica das VAs seja aplicada em contextos compulsórios, como avaliações para o trânsito, porte de arma de fogo, contextos cirúrgicos, seleção e recrutamento, haja vista que são situações em que existe uma tendência dos indivíduos em não se apresentar de forma totalmente honesta, bem como ter uma variabilidade maior em relação às questões sociodemográficas, como nível de renda, escolaridade etc.

O segundo ponto para estudos futuros se dá no que se conhece sobre os processos de resposta quando falamos em erros do tipo sistemático. Aqui não foi possível avaliar quais os processos mentais e como eles se dão na utilização de uma escala de resposta, em especial de uma técnica para mensurar este tipo de erro sistemático. Estudos desta natureza podem nos

ajudar a compreender, a partir do uso de Vinhetas-âncora, como um mesmo atributo pode gerar diferentes percepções e ativar processos cognitivos necessários para se responder a uma situação exposta.

O terceiro ponto se refere a compreender como o erro sistemático conhecido como funcionamento diferencial da pessoa (DPF) se relaciona com os estilos de resposta. Como exposto, a técnica das VAs foi proposta, inicialmente, como uma forma de compreender os diferentes usos de escalas em diferentes populações e a isso foi denominado funcionamento diferencial do item (DIF) pelos primeiros estudos, mas, como exposto na Introdução, optamos por compreender o que os estudos de Vinhetas-âncora chamam assim como DPF. Contudo, tal termo é utilizado na perspectiva da Teoria de Resposta ao Item (TRI), enquanto outras perspectivas mencionam o termo de erros sistemáticos, em especial estilos de respostas. Assim, estudos que visem compreender possíveis nuances entre estilos de respostas e DPF, com esta técnica das VAs, pode ser útil em avançar no campo tanto metodológico quanto teórico da medida.

Um possível quarto ponto, em termos de agenda de pesquisa, seria a possibilidade de aumentar o banco de situações das Vinhetas-âncora para testar como a variância, em mais situações hipotéticas, se comportam. Isso pode nos dar uma visão mais clara sobre as estimativas de precisão, de forma a melhorar essa característica. Ainda, isso também possibilitaria uma análise multinível para comparar como o Coeficiente de Correlação Intraclasse se apresenta a partir dos modelos testados e com o controle das variáveis em questão.

Salienta-se que mais estudos acerca do uso de Vinhetas-âncora para mensuração e controle do estilo de resposta de viés de grupo sejam necessários, haja vista que a literatura tem tido diversos pontos divergentes sobre o quanto as VAs podem beneficiar instrumentos de autorrelato aplicados em avaliações de larga-escala e transculturais neste quesito. Por isto, recomenda-se que demais estilos de resposta, como medidas extremas, tendência central e

resposta socialmente desejável, que também são fonte de erros do tipo sistemático, possam ser estudados a partir do uso da técnica de Vinhetas-âncora.

Diante do exposto, espera-se que este estudo consiga servir de norte nos seus avanços aqui discutidos, bem como ajude em futuras pesquisas com a agenda de pesquisa proposta. Apesar de se ter um método ou técnica que controle ou mensure por completo erros do tipo sistemático seja quase utópico, conseguir avançar cada vez mais nesse sentido contribui fortemente para avaliações mais fidedignas. Com resultados mais precisos, o impacto em políticas públicas e tomadas de decisões, em diferentes níveis, é diretamente beneficiada e possibilita uma transformação social no qual a ciência aqui produzida se propõe.

Referências

- Abrahams, L., Pancorbo, G., Primi, R., Santos, D., Kyllonen, P., John, O. P., & De Fruyt, F. (2019). Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts. *Psychological Assessment, 31*(4), 460-473. <https://dx.doi.org/10.1037/pas0000591>.
- Adams, R., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychologic Measurement 21*(1), 1–23. <https://dx.doi.org/10.1177/0146621697211001>
- Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed mode panel studies. *Social Science Research, 42*, 957-970. <https://dx.doi.org/10.1016/j.ssresearch.2013.01.002>
- Andrich, D., Humphry, S. M., & Marais, I. (2012). Quantifying local, response dependence between two polytomous items using the Rasch model. *Applied Psychological Measurement, 36* (4), 309-324. <https://dx.doi.org/10.1177/0146621612441858>
- Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behav. Res. Methods 52*, 2489–2505. <https://dx.doi.org/10.3758/s13428-020-01401-8>
- Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., Egeland, T. (2018). The failing measurement of attitudes: How semantic determinants of individual survey responses come to replace measures of attitude strength. *Behavior Research Methods, 50*, 2345–2365. <https://doi.org/10.3758/s13428-017-0999-y>
- Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences, 40*, 1235-1245. <https://dx.doi.org/10.1016/j.paid.2005.10.018>

- Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*, 491-509. <https://dx.doi.org/10.1086/268845>
- Batchelor, J. H., & Miao, C. (2016). Extreme response style: A meta-analysis. *Journal of Organizational Psychology*, *16*, (2), 50-62. Retrieved from https://www.researchgate.net/publication/316820164_Extreme_Response_Style_A_Meta-Analysis
- Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665-678. <https://dx.doi.org/10.1037/a0028111>
- Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods*, *22*, 69-83. <https://dx.doi.org/10.1037/met0000106>
- Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical & Statistical Psychology*, *70*, 159-181. <https://dx.doi.org/10.1111/bmsp.12086>
- Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. New York: John Wiley & Sons.
- Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, *33*, 335-352. <https://dx.doi.org/10.1177/0146621608329891>
- Bolt, D. M., Lu, Y., & Kim, J.-S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, *19*, 528-541. <https://dx.doi.org/10.1037/met0000016>
- Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, *71* (5), 814-833. <https://dx.doi.org/10.1177/0013164410388411>

- Borkenau, P., & Ostendorf, F. (2008). NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI). Manual (2. Auflage). Göttingen: Hogrefe.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71* (3), 460-502. <https://dx.doi.org/10.1177/0013164410375112>
- Cabooter, E. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures*. Ghent University, Unpublished Doctoral Dissertation, Ghent, Belgium. Retrieved from <https://biblio.ugent.be/publication/4333765/file/4427719>
- Carroll, J. (1963). A model of school learning. *The Teachers College Record, 64*(8), 723-723.
- Casey, M. M., & Tryon, W. W. (2001). Validating a double-press method for computer administration of personality inventory items. *Psychological Assessment, 13*, 521-530. <https://dx.doi.org/10.1037/1040-3590.13.4.521>
- Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272–295. <https://doi.org/10.3102/0034654310362998>
- Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology, 33* (6), 401-415. <https://dx.doi.org/10.1037/h0054677>
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475-494. <https://dx.doi.org/10.1177/001316444600600405>
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10* (1), 3-31. <https://dx.doi.org/10.1177/001316445001000101>

- Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*(1), 21-50. https://dx.doi.org/10.1207/s15327752jpa6401_2
- Costa, P. T., Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, *55*, 258-265. <https://dx.doi.org/10.1037/0022-3514.55.2.258>
- Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119-130. <https://dx.doi.org/10.1016/j.jrp.2015.05.004>
- De Fruyt, F., Wille, B., & John, O. P. (2015). Employability in the 21st Century: Complex (Interactive) problem solving and other essential skills. *Industrial and Organizational Psychology*, *8*, 276–281. <https://dx.doi.org/10.1017/iop.2015.33>.
- DeMars, C. E. (2018). Classical Test Theory and Item Response Theory. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale, and test development* (pp. 413-443). West Sussex: John Wiley & Sons.
- De Raad, B., & Perugini, M. (Eds.). (2002). *Big five assessment*. Seattle, WA: Hogrefe & Huber.
- De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality*, *10*, 303–336.
- Digman, J. M. (2002). Historical antecedents of the Five-Factor Model. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the Five-Factor Model of Personality* (pp. 17- 22). Washington, DC: American Psychological Association.

- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. <https://dx.doi.org/10.1111/j.1467-8624.2010.01564.x>
- Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, 16, 20-30. <https://dx.doi.org/10.1027//1015-5759.16.1.20>
- Estefania, Caicedo Cavagnis, & Zalazar-Jaime, Mauricio Federico. (2018). Entrevistas cognitivas: revisión, directrices de uso y aplicación en investigaciones psicológicas. *Avaliação Psicológica*, 17(3), 362-370. <https://dx.doi.org/10.15689/ap.2018.1703.14883.09>
- Falk, C., & Cai, L. (2016). A flexible full-information approach to modeling of response styles. *Psychological Methods*, 21(3), 328-347. <https://dx.doi.org/10.1037/met0000059>
- Fekken, G. C., & Holden, R. R. (1994). The construct validity of differential response latencies in structured personality tests. *Canadian Journal of Behavioural Science*, 26, 104-120. <https://dx.doi.org/10.1037/0008-400X.26.1.104>
- Ferrando, P. J., & Lorenzo-Seva, U. (2007). An item-response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement*, 31, 525-543. <https://dx.doi.org/10.1177/0146621606295197>
- Ferrando, P. J. & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology*, 63, 427–448. <https://dx.doi.org/10.1348/000711009X470740>

- Ferrando, P. J., Morales-Vives, F., & Lorenzo-Seva, U. (2016). Assessing and controlling acquiescent responding when acquiescence and content are related: A comprehensive factor-analytic approach. *Structural Equation Modeling*, 23, 713–725. <https://dx.doi.org/10.1080/10705511.2016.1185723>
- Fischer, R. (2004). Standardization to account for cross-cultural response bias. *Journal of Cross-Cultural Psychology*, 35, 263–282. <https://dx.doi.org/10.1177/0022022104264122>
- Fowler, F. J.; Lloyd, S. J.; Cosenza, C. A. & Wilson, I. B. (2014). Coding Cognitive Interview: an approach to enhancing the value of cognitive testing for survey question evaluation. *Field Methods*, 28(1) 1-18. <https://dx.doi.org/10.1177/1525822X14549921>.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and Social Psychology* (pp. 141-165). Beverly Hills, CA: Sage.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1), 26-42. <https://dx.doi.org/10.1037/1040-3590.4.1.26>.
- Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models. *Journal of Psychology*, 216(2), 89-101. <https://dx.doi.org/10.1027/0044-3409.216.2.89>
- Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, 23, 1440-1465. <https://dx.doi.org/10.3758/s13423-016-1025-6>
- Henninger, M. (2019). A novel varying threshold IRT approach to accounting for response styles. *Manuscript Submitted for Publication to the Journal of Educational Measurement*.

- Henninger, M., & Meiser, T. (2019a). Different approaches to modeling response styles in Divide-by-Total IRT models (Part I): A model integration. *Invited Revision Submitted to Psychological Methods*.
- Henninger, M., & Meiser, T. (2019b). Different approaches to modeling response styles in Divide-by-Total IRT models (Part II): *Applications and novel extensions*. *Invited Revision Submitted to Psychological Methods*.
- Henninger, M., & Plieninger, H. (2019). Different styles, different times: How response times can inform our knowledge about the response process in rating scales. *Revision Invited by Assessment*.
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behavior Research Methods*, 39, 101-117. <https://dx.doi.org/10.3758/BF03192848>
- Hui, C. H., & Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49, 253_260. <https://dx.doi.org/10.1086/268918>
- Hopkins, D., & King, G. (2010). Improving anchoring vignettes designing surveys to correct interpersonal incomparability. *The Public Opinion Quarterly*, 74(2), 201-222. Retrieved from: <http://www.jstor.org/stable/40660640>
- Johanson, G., & Alsmadi, A. (2002). Differential person functioning. *Educational and Psychological Measurement*, 62(3), 435-443
- Instituto Ayrton SENNA [IAS] (2020). Quem somos. <https://www.institutoayrtonSENNA.org.br/pt-br/quem-somos.html>
- Johanson, G. A., & Osborn, C. J. (2004). Acquiescence as differential person functioning. *Assessment and Evaluation in Higher Education*, 29(5), 535-548. <https://dx.doi.org/10.1080/02602930410001689126>

- John, O. P., & De Fruyt, F. (2015). *Framework for the Longitudinal Study of Social and Emotional Skills in Cities*. Paris: OECD
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York: Guilford Press
- John, O. P., & Robins, R. W. (1993). Gordon Allport: Father and critic of the five-factor model. In K. H. Craik, R. Hogan, & R. N. Wolfe (Eds.), *Fifty years of personality psychology* (pp. 215–236). New York: Plenum Press.
- Kaufman, J. C. (2012). Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 298-308. <https://dx.doi.org/10.1037/a0029751>
- Karami, H. (2012). The development and validation of a bilingual version of the vocabulary size test. *RELC Journal*, 43(1). <https://dx.doi.org/10.1177/0033688212439359>
- King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(01), 191-207.
- Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, 77, 379-386. <https://dx.doi.org/10.1037/0022-3514.77.2.379>
- Kyllonen, P. C., & Bertling, J. P. (2014). *Draft Report: Anchoring Vignettes Reduce Bias in Noncognitive Rating Scale Responses*. Princeton, New Jersey: ETS/OECD.
- Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). Personality, Motivation, and College Readiness: A prospectus for assessment and development. *ETS Research*

- Report Series*, 1–48. Princeton, NJ: Educational Testing Service. <https://dx.doi.org/10.1002/ets2.12004>
- Lelkes, Y., & Weiss R. (2015). Much ado about acquiescence: The relative validity and reliability of construct-specific and agree–disagree questions. *Research & Politics*, July–September, 1-8. <https://dx.doi.org/10.1177/2053168015604173>
- Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Struct. Equ. Modeling* 14, 26–47.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13, 203-229. <https://dx.doi.org/10.1037/a0012869>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174. <https://dx.doi.org/10.1007/BF02296272>
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11, 344-362. <https://dx.doi.org/10.1037/1082-989X.11.4>
- Marsh, H. W. (1987). The Big-Fish-Little-Pound effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280-295.
- Marsh, H. W., & Parker, J. W. (1984). Determinants of students' self-concept: Is it better to be a relatively large fish in a small pound even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1), 213-231
- Meijer, R. R., & Tendeiro, J. N. (2018). Unidimensional item response theory. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale, and test development* (pp. 413-443). West Sussex: John Wiley & Sons.

- McCrae, R. R. (2018). Method biases in single-source personality assessments. *Psychological Assessments*. Advance online publication. <https://dx.doi.org/10.1037/pas0000566>
- Messick, S. (1980). Test validity and ethics of assessment. *American Psychologist* 35(11), 1012-1027. <https://dx.doi.org/10.0003-066X/80/3511-1012>
- Moors, G. (2003). Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality and Quantity*, 37, 277-302. <https://dx.doi.org/10.1023/A:1024472110002>
- Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review*, 20, 303-320. <https://dx.doi.org/10.1093/esr/jch026>
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity*, 42 (6), 779-794. <https://dx.doi.org/10.1007/s11135-006-9067-x>
- Mõttus, R., Allik, J., Realo, A., Pullmann, H., Rossier, J., Zecca, G., . . . Tseung, C. N. (2012). Comparability of self-reported conscientiousness across 21 countries. *European Journal of Personality*, 26(3), 303-317. <https://dx.doi.org/10.1002/per.840>.
- Organization for Economic Co-operation Development. (2015). *Education at a Glance 2015: OECD Indicators [OECD]*. OECD Publishing. <https://dx.doi.org/10.1787/eag-2015-en>
- Organization for Economic Co-operation Development. (2018). *Education at a Glance 2018: OECD Indicators [OECD]*. OECD Publishing. <https://dx.doi.org/10.1787/eag-2018-en>
- Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. P. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. *Review of Research in Education*, 40(1), 644–681. <https://dx.doi.org/10.3102/0091732X16673595>

- Padilla, J. L. & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136-144.
<https://dx.doi.org/10.7334/psicothema2013.259>.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17-59). <https://dx.doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Pasquali, L., & Primi, R. (2003). Fundamentos da Teoria da Resposta ao Item - TRI. *Avaliação Psicológica*, 2, 99-110.
- Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structure from personality-trait descriptors. *Journal of Personality and Social Psychology*, 57(3), 552-567
- Plieninger, H. (2017). Mountain or molehill? A simulation study on the impact of response styles. *Educational and Psychological Measurement*, 77, 32-53.
<https://dx.doi.org/10.1177/0013164416636655>
- Plieninger, H., & Heck, D. W. (2018). A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research*, 53, 633-654. <https://dx.doi.org/10.1080/00273171.2018.1469966>
- Plieninger, H., Henninger, M., & Meiser, T. (2019). An experimental comparison of the effect of different response formats on response styles. Manuscript submitted for publication.
- Primi, R.; Zanon, C.; Santos, D.; De Fruyt, F.; John, O. P. (2016a). Can Anchoring Vignettes Make Adolescent Self-reports of Social-emotional Personality Traits More Reliable, Discriminant, and Criterion-valid? *European Journal of Psychological Assessment*, 32(1), 39-51. <https://dx.doi.org/10.1027/1015-5759/a000336>
- Primi, R.; Santos, D.; John, O. P.; De Fruyt, F. (2021). *SENNA Inventory for the Assessment of Social and Emotional Skills: Technical Manual*. Instituto Ayrton Senna.
<https://doi.org/10.31234/osf.io/byvpr>

- Primi, R; Santos, D.; De Fruyt, F. & John, O. P. (2016b). Development of an inventory assessing social and emotional skills in brazilian youth. *European Journal of Psychological Assessment*, 32(1), 5-16. <https://dx.doi.org/10.1027/1015-5759/a000343>.
- Primi, R; Santos, D.; De Fruyt, F. & John, O. P. (2019). Comparison of classical and modern methods for measuring and correcting for acquiescence. *British Journal of Mathematical and Statistical Psychology, Special Issue*. <https://dx.doi.org/10.1111/bmsp.12168>
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), p.322-338.
- R Development Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna: Austria. Retrieved from <http://www.R-project.org>.
- Revelle, W. (2021). *psych: Procedures for Personality and Psychological Research. R package version 2.2.0*. <https://personality-project.org/r/psych-manual.pdf>
- Roberts, B.W., Walton, K.E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1-25.
- Santos, D. & Primi, R. (2014). *Desenvolvimento socioemocional e aprendizado escolar: uma proposta de mensuração para apoiar políticas públicas*. Relatório sobre resultados preliminares do projeto de medição de competências socioemocionais no Rio de Janeiro. São Paulo: OCDE, SEEDUC, Instituto Ayrton SENNA.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99-144). New York, NY, US: The Guilford Press.

- Silva, A. S., Campos-Silva, W. L., Gouvea, M. A., & Farina, M. C. (2018). Vignettes: Uma técnica de coleta de dados para lidar com o funcionamento diferencial dos itens em levantamento de dados. *Brazilian Business Review*, *16*(1), 16-31. <https://dx.doi.org/10.15728/bbr.2019.16.1.2>
- Soto, C.J., & John, O.P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, *113*(1), 117-143. <https://dx.doi.org/10.1037/pspp0000096>
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, *94*, 718–737. <https://dx.doi.org/10.1037/0022-3514.94.4.718>.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*(2), 330-348. <https://dx.doi.org/10.1037/a0021717>
- Steinmann, I., Strietholt, R., and Braeken, J. (2021). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychol. Methods* *1*, 36. <https://dx.doi.org/10.1037/met0000392>
- Ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research*, *34*, 89-102
- Thissen, D., Steinberg, L., & Wainer, H. (1993). *Detection of differential item functioning using the parameters of item response models*. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (p. 67–113). Lawrence Erlbaum Associates, Inc.

- Tracey, T. J. G. (2016). A note on socially desirable responding. *Journal of Counseling Psychology, 63*(2), 224-232. <https://dx.doi.org/10.1037/cou0000135>
- van der Linden, W. J., Klein Entink, R. H., & Fox, J.-P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement, 34*, 327-347. <https://dx.doi.org/10.1177/0146621609349800>
- van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika, 68* (2), 251-265. <https://dx.doi.org/10.1007/BF02294800>
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25*, 195-217. <https://dx.doi.org/10.1093/ijpor/eds021>
- van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*, 346-360. <https://dx.doi.org/10.1177/0022022104264126>
- Von Davier, M., Shin, H. J., Khorramdel, L., Stankov, L. (2017). The effects of vignette scoring on reliability and validity of self-reports. *Applied Psychological Measurement, 42*(4). <https://dx.doi.org/10.1177/0146621617730389>
- Vonkova, H., Bendl, S., & Papajoanu, O. (2017) How students report dishonest behavior in school: Self-assessment and anchoring vignettes, *The Journal of Experimental Education, 85*(1), 36-53. <https://dx.doi.org/10.1080/00220973.2015.1094438>
- Wand, J., King, G., & Lau, O. (2011). Anchors: Software for anchoring vignette data. *Journal of Statistical Software*. Forthcoming, URL <http://www.jstatsoft.org>.
- Wang, W.-C., & Wu, S.-L. (2011). The random-effect generalized rating scale model. *Journal of Educational Measurement, 48*, 441-456. <https://dx.doi.org/10.1111/j.1745-3984.2011.00154.x>

- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 236-247. <https://dx.doi.org/10.1016/j.ijresmar.2010.02.004>
- Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, 34, 105-121. <https://dx.doi.org/10.1177/0146621609338593>
- Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological Methods*, 15, 96-110. <https://dx.doi.org/10.1037/a0018721>
- Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, 36 (3), 409-422. <https://dx.doi.org/10.1007/s11747-007-0077-6>
- Wetzel, E., & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, 33, 352-364. <https://dx.doi.org/10.1027/1015-5759/a000291>
- Wetzel, E., Carstensen, C. H., & Böhnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality*, 47, 178-189. <https://dx.doi.org/10.1016/j.jrp.2012.10.010>
- Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 23(3), 279-291. <https://dx.doi.org/10.1177/1073191115583714>