Nemvus Produções, Acesso Aberto | www.nemvus.com
8 de Março, 2022

Confiabilidade Entre Avaliadores do Código Vultológico

Juan E. Sandoval

Traduzido por Hugo G. P. Rocha

Abstract: The reliability of the Cognitive Type Vultology Code 3 (CTVC3), as a visual analysis instrument, is evaluated through a study involving the classification of random samples by ten raters to test the level of independent convergence among them. During the course of six weeks, ten students were trained in the methodology of the instrument. Towards the end of the six weeks, a final exam containing ten randomly selected people was administered to the raters privately, to assess their degree of independent agreement with the instructor and with each other.

Palavras-chave: *Expressões Faciais, Análise Facial, Linguagem Corporal, Vultologia, Junguiano, Carl Jung, Cognição Incorporada, Confiabilidade Entre Avaliadores*

1. Introdução

O CTVC3 é um instrumento de análise visual recém-desenvolvido, usado para classificação através da avaliação combinada de expressões faciais, maneirismos corporais e tom de voz. Esses sinais visuais parecem estar estatisticamente conectados a elementos da psicologia, como dados ocupacionais,[1] permitindo-nos prever elementos da personalidade de uma pessoa através de suas expressões corporais. No entanto, nenhum estudo até o momento empreendeu a análise da confiabilidade do instrumento na classificação de indivíduos da mesma forma sob condições controladas. Questões sobre a eficácia do novo instrumento devem ser abordadas antes que ele possa ser implementado em casos de uso mais amplos. Este estudo tem como objetivo abordar a questão seguindo os resultados do primeiro curso de prática sobre o CTVC3, que visava ensinar a metodologia a uma turma. Doze alunos foram inscritos no programa, liderados por Juan E. Sandoval como instrutor. Dois alunos não puderam participar do exame final. Dos dez estudantes que participaram do exame final, quatro (estudantes 4, 5, 6, 7) tiveram exposição significativa anterior à metodologia, enquanto os seis estudantes restantes tiveram pouca ou nenhuma exposição prévia ao CTVC3.

2. Período de Treinamento

O período de treinamento de seis semanas, entre janeiro e fevereiro de 2022, envolveu os seguintes itens gerais:

1. **Uma série de palestras em vídeo sobre a metodologia**, explicando as origens do instrumento, sua aplicação adequada em um ambiente clínico, bem como a teoria por trás de seu uso.
2. **Uma série de quatro testes online**, contendo aproximadamente 20 perguntas cada. Os testes utilizaram GIFs animados para demonstrar sinais visuais e deram respostas de múltipla escolha sobre o que está sendo exibido no GIF animado fornecido.

Figura 1

	Estudante 1	Estudante 2	Estudante 3	Estudante 4	Estudante 5	Estudante 6	Estudante 7	Estudante 8	Estudante 9	Estudante 10	**Resultados do Instrutor**	**Consenso do Grupo**
1. Jimmy Donaldson	SeTi l-l-	NeFi l-l-	SeFi l-l-	SeFi l---	SeFi l---	SeFi l-l-	NeTi l-l-	TiNe	SeFi l---	TeNi	**SeFi l-l-**	**SeFi**
2. Ray Dalio	TeSi ll--	TeSi ll--	FeSi l---	FeSi ll--	FeSi ll--	FeSi l---	TeSi l---	TeSi	TeSi l---	NeTi	**FeSi l---**	**FeSi**
3. Colin Powell	TeNi ll--	SiFe ll--	FeSi ll--	FeNi ll--	FeNi ll--	FeNi ll--	FeSi ll--	FeNi	TeNi l---	NiTe	**FeNi ll--**	**FeNi**
4. Usain Bolt	SeTi l---	SeTi l--l	SeTi l---	SeTi l---	SeTi l---	SeTi l-l-	SeTi l--l	SeTi	SeTi l---	SeTi	**SeTi l---**	**SeTi**
5. Lee Ji-eun	TiSe l---	TiSe ll--	TiNe ll-l	TiSe l---	TiSe l---	FiNe l---	TiNe ll--	FiSe	TiSe l---	FiNe	**TiSe ll--**	**TiSe**
6. Becky G	SeFi l-l-	SeFi l---	NeTi l---	SeFi l---	SeFi l-l-	SeFi l---	SeFi l-l-	SeFi	NeFi l---	SeFi	**SeFi l---**	**SeFi**
7. Ryan Seacrest	FeNi l-l-	FeSi l-l-	SeTi l-l-	FeNi l-l-	FeNi l---	FeNi l---	FeSi l-l-	NeTi	FeSi l---	SeTi	**FeNi l-l-**	**FeNi**
8. Eva Longoria	FeNi ll--	FeNi l-ll	FeSi l-ll	FeNi l-l-	FeNi l---	FeNi l---	SeTi l-l-	SeFi	FeNi l---	FeSi	**FeNi l-l-**	**FeNi**
9. Janet Conrad	SiTe lll-	TeSi l---	FeSi l---	TeSi ll--	FeSi l---	TeSi l---	TeSi l---	FeSi	FiNe l---	SiTe	**TeSi l---**	**TeSi**
10. Sabrina Carpenter	SeFi l-l-	FeNi l-l-	NeTi l-l-	SeFi l-l-	NeFi l---	SeFi l---	SeFi l-l-	SeFi	FiSe l---	FeNi	**NeFi l---**	**SeFi**
Pontos / Instrutor	35/40	33/40	32/40	39/40	39/40	37/40	32/40	29/40	33/40	28/40	N/A	N/A
Pontos / Consenso	36/40	34/40	31/40	40/40	38/40	38/40	33/40	30/40	34/40	29/40	**39/40**	N/A
Perc. / Instrutor	**87.50%**	**82.50%**	**80%**	**97.50%**	**97.50%**	**92.50%**	**80%**	**72.50%**	**82.50%**	**70%**	N/A	N/A
Perc. / Consenso	**90%**	**85%**	**77.50%**	**100%**	**95%**	**95%**	**82.50%**	**75%**	**85%**	**72.50%**	**97.50%**	N/A

Dada a natureza do teste composto por quatro questões dicotômicas, uma pontuação geral de 50% (20/40) não seria melhor do que o acaso. No entanto, todos os dez avaliadores pontuaram significativamente acima do acaso, com pontuações variando de 28/40 a 39/40 em relação aos resultados do instrutor, e entre 29/40 a 40/40 em relação aos resultados do consenso. Curiosamente, os resultados do instrutor e os resultados do consenso concordaram entre si em 39/40 pontos (97,5%). As Figuras 2 e 3 representam melhor esses dois resultados em gráficos. A pontuação média entre todos os dez avaliadores foi de 337/400 (84,25%).

Comentário do Instrutor

Alunos 8 e 10, enviaram seus resultados finais dentro de minutos do prazo final e não puderam dedicar tanto tempo quanto os outros participantes do exame devido a circunstâncias pessoais. Eles também parecem ter pontuado mais baixo do que outros alunos, sugerindo que maiores quantidades de investimento de tempo na metodologia de digitação se correlacionam positivamente com maiores acordos entre avaliadores. Isso está de acordo com as expectativas de uma ferramenta diagnóstica funcional. Dos quatro alunos que passaram no limite do instrutor para certificação de 90%+ (alunos 1, 4, 5, 6), apenas o aluno 1 não fazia parte do grupo previamente familiarizado. No entanto, eles também estavam entre os mais engajados na prática do método durante o curso e foram os primeiros a enviar seus resultados. Isso sugere que o curso de seis semanas foi capaz de levar um indivíduo a uma concordância de 90% entre avaliadores previamente treinados quando a participação era ideal.

3. **Uma série de quatro reuniões em grupo online**, aproximadamente uma semana de intervalo, para revisar os resultados do teste e responder às perguntas dos alunos.
4. **Uma tarefa pessoal**, envolvendo a criação de um relatório de vultologia para uma amostra de celebridades aleatórias.
5. **Uma prova final**, envolvendo classificar 10 indivíduos selecionados aleatoriamente sob condições cegas.

Ainda assim, apesar dessas disposições, o curso intensivo de seis semanas não foi capaz de ensinar todo o material relacionado ao uso do instrumento CTVC3, cobrindo apenas os elementos introdutórios à prática. No entanto, uma quantidade suficiente dos fundamentos foi comunicada para avaliar a adequação geral do método para educação.

3. Configuração do Exame

O CTVC3 é composto por 71 sinais visuais e auditivos. Sessenta e quatro desses sinais foram utilizados no exame para classificar as pessoas em quatro categorias principais cada:

- Proativo vs Reativo *(corr. E vs I)*
- Condutor vs Revisor *(corr. Je-Pi vs Pe-Ji)*
- Franco vs Ponderado (*corr. Fi-Te vs Ti-Fe)*
- Aterrado vs Suspendido *(corr. Ni-Se vs Ne-Si)*

Essas quatro categorias dicotômicas representam o número mínimo de pontos de dados necessários para classificação do tipo em um dos dezesseis resultados possíveis. Consequentemente, cada digitação foi construída com base em quatro pontos, um para cada dicotomia correspondente. Quando multiplicado por dez amostras, isso resultou em uma pontuação total de 40 pontos possíveis para o exame final. Uma pontuação mínima para certificação na metodologia foi estabelecida em 36/40 pontos (90%) seja em relação ao instrutor ou ao consenso do grupo.

As dez amostras usadas para o estudo foram selecionadas aleatoriamente usando uma ferramenta online que escolheu vídeos do YouTube através de um algoritmo. No entanto, foi exercida alguma discrição na remoção de vídeos sem indivíduos em destaque, indivíduos com muito pouco material utilizável online, ou indivíduos que já foram digitados anteriormente. Exceto por esses requisitos, quaisquer indivíduos aleatórios com material de vídeo suficiente disponível no YouTube para digitação foram aprovados até que dez amostras desse tipo fossem adquiridas para o exame.

A lista de dez amostras foi fornecida ao mesmo tempo aos dez avaliadores através de correspondência privada. Foi instruído aos avaliadores que não se comunicassem entre si ou com mais ninguém sobre as amostras durante a duração do teste. Eles foram instruídos a utilizar a ferramenta Codificador online para inserir seus sinais visuais e exportar um relatório em PDF com seus resultados para cada amostra, e então enviar de volta os dez PDFs juntos para o instrutor antes do prazo final. Dado o longo tempo necessário para completar o teste, e para ajudar a acomodar horários díspares, a janela de teste foi definida para 9 dias, cobrindo dois finais de semana. Durante esse período, o instrutor também realizou a análise das dez novas amostras em paralelo com os alunos.

4. Resultados do Exame

Dois métodos de determinar os resultados foram considerados simultaneamente. O primeiro método envolveu acordo com os próprios resultados do instrutor para as dez amostras randomizadas. No entanto, porque o instrutor também é um praticante e avaliador usando a ferramenta, ele não está isento de erro instrumental potencial. Para combater isso, um segundo método de medir os resultados foi usado, envolvendo acordo com o consenso do grupo. A pontuação do consenso foi calculada somando os pontos dados às quatro dicotomias de uma amostra, por todos os avaliadores, para chegar ao tipo correspondente. A Figura 1 mostra os resultados para ambas essas medidas, usando uma escala de 40 pontos.

Figura 2

Resultados Comparados ao Instrutor

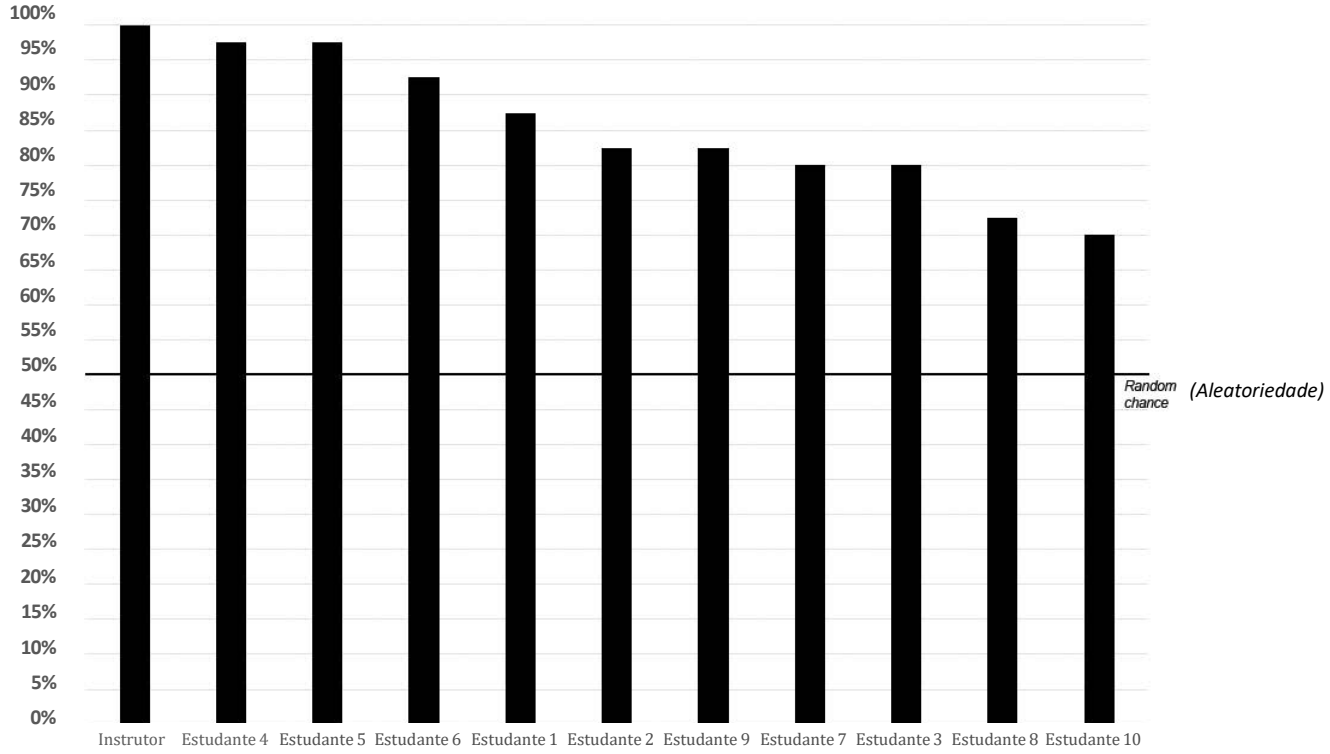
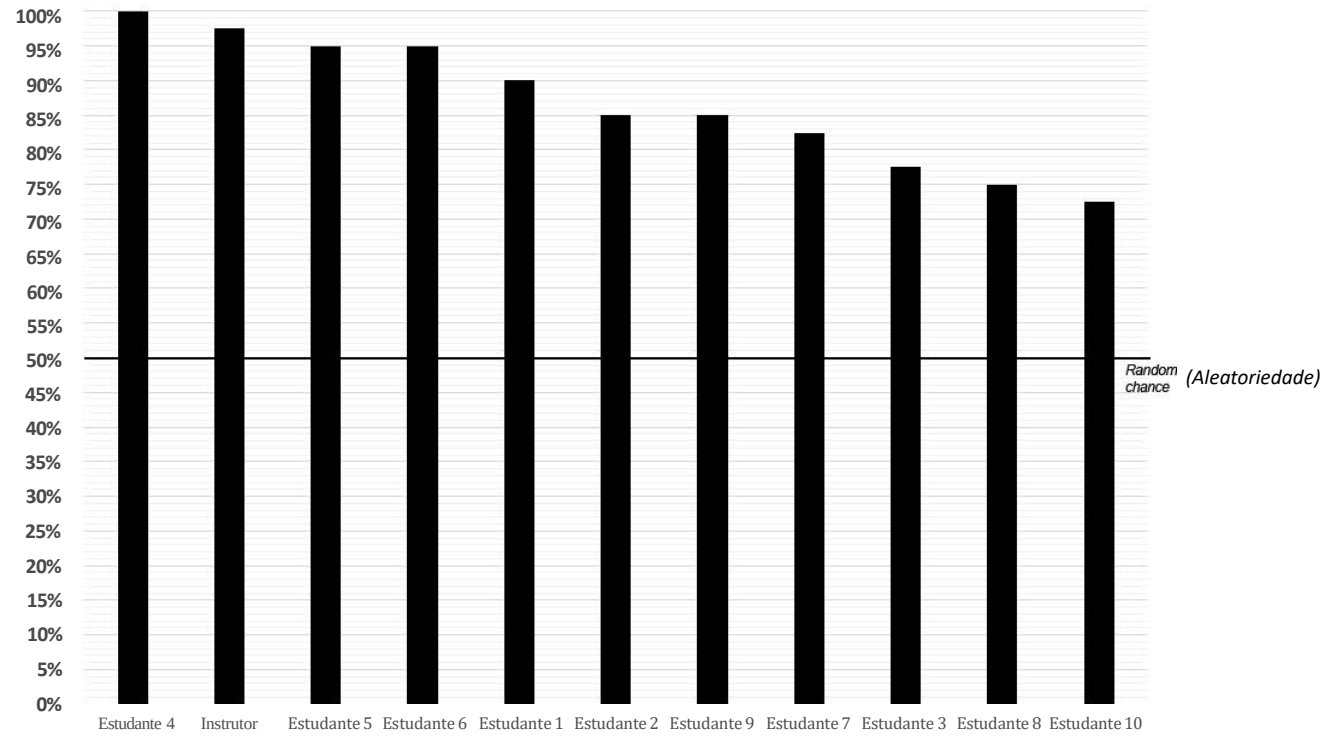


Figura 3

Resultados Comparados ao Consenso



5. Significância Estatística

Para estimar a significância estatística entre os acordos entre avaliadores, foi realizada uma análise usando o valor de Kappa de Cohen, que ajusta para possível concordância devido ao acaso de acordo com a seguinte fórmula:

$$K \quad \frac{P_{concordância} - P_{acaso}}{1 - P_{acaso}}$$

$P_{concordância}$ = Proporção de julgamentos nos quais os juízes concordam

P_{acaso} = Proporção de julgamentos nos quais a concordância seria esperada devido ao acaso

A probabilidade de concordância P_c é mostrada nas Figuras 2 e 3, enquanto a probabilidade de chance P_a é calculada em .5 (ou 50%). Os resultados são então listados nas Figuras 4 e 5 para os dois métodos de pontuação correspondentes, ao lado da escala de significância de acordo com a interpretação padrão dos valores de Kappa por Altman (1999) & Landis JR (1977).

Figura 4

Valores Kappa Relativos ao Instrutor

	Kappa	Significância
Estudante 1	0.75	Substancial
Estudante 2	0.6	Moderada
Estudante 3	0.6	Moderada
Estudante 4	0.95	Quase Perfeita
Estudante 5	0.95	Quase Perfeita
Estudante 6	0.85	Quase Perfeita
Estudante 7	0.6	Moderada
Estudante 8	0.45	Moderada
Estudante 9	0.6	Moderada
Estudante 10	0.4	Razoável

Figura 5

Valores Kappa Relativos ao Consenso

	Kappa	Significância
Estudante 1	0.8	Quase Perfeita
Estudante 2	0.65	Substancial
Estudante 3	0.55	Moderada
Estudante 4	1	Quase Perfeita
Estudante 5	0.9	Quase Perfeita
Estudante 6	0.9	Quase Perfeita
Estudante 7	0.65	Substancial
Estudante 8	0.5	Moderada
Estudante 9	0.65	Substancial
Estudante 10	0.45	Moderada

Depois dos resultados terem sido calculados independentemente, ocorreu uma reunião final entre o instrutor e colegas em que os resultados foram discutidos mutuamente. Durante o discurso subsequente, a única diferença entre as pontuações do instrutor e as pontuações de consenso foi examinada, e descobriu-se que a pontuação de consenso estava mais adequadamente alinhada ao CTVC3. O instrutor cedeu à avaliação de consenso da amostra 10 como SeFi, não NeFi, enquanto todos os outros pontos permaneceram inalterados. O cálculo de consenso mostrou-se uma avaliação ligeiramente melhor das pontuações gerais da amostra, mas apenas retrospectivamente após o término do exame cego. No entanto, ambas as formas de pontuação (antes e após o discurso em grupo) são retidas neste estudo para demonstrar avaliações cegas e não cegas.

6. Conclusões

O método CTVC3 de avaliação de expressões visuais parece ser capaz de atender aos padrões necessários de confiabilidade entre avaliadores para uma aplicação mais ampla em testes clínicos e acadêmicos. Um curso de treinamento de seis semanas é capaz de produzir um acordo interavaliadores *"substancial"* naqueles que começam com pouco conhecimento (como visto com o Estudante 1), e uma educação prolongada é capaz de produzir um acordo *"Quase Perfeito"* (como visto com os Estudantes 4, 5, 6).

1. Sandoval, J. e Hershkoviz, H. “Diferenças Ocupacionais Entre Tipos Vultológicos", Nemvus Produções, 2021, https://nemvus.com/article/127/
2. https://perchance.org/youtube-video
3. https://vultology.com/codifier/