

Sobre a medição da dinâmica da formação da Identidade de Gênero – e de suas características em relação a testes psicométricos clássicos.

por Wal Torres, MSc., Ph.D.

Gendercare Gender Clinic

Copyright © 2007, Torres. Todos os direitos reservados.

Abstract:

O principal objetivo de nosso trabalho, não tem sido o de estudar as variâncias de gênero – GV nem a formação da identidade de gênero – GI, quanto a sua etiologia. Pelo contrário, nosso maior objetivo tem sido a medição da dinâmica de formação da identidade de gênero, o que permite a comparação de indivíduos (“pacientes”) com estados típicos - não se considerando a etiologia da formação e muito menos procurando explicações reducionistas tipo uma causa – um efeito. Essa quantificação é possível se considerarmos assinaturas dinâmicas típicas – um método já usado em outras áreas do conhecimento – exemplos em física, astrofísica e em neurobiologia – considerando que a identidade de gênero se forma a partir de muitas causas que interagem entre si – gerando múltiplos efeitos. Experimentalmente conseguimos levantar dados para definir assinaturas dinâmicas típicas para as famílias de estados definidos pelos SOC 6 da HBGDA/WPATH – podendo então caracterizar e comparar assinaturas. Essa avaliação e comparação são feitas de forma privilegiada usando-se computadores e levantando-se os dados on-line. Quando alguém evidencia a possibilidade de sofrer de um transtorno psiquiátrico real que possa levar a um verdadeiro transtorno de identidade essa pessoa é classificada como GIDNOS – e pode necessitar de uma avaliação mais detalhada e um acompanhamento especial – mas esses casos são raros. Ênfase é dada aqui às diferenças conceituais básicas entre medições psicométricas clássicas e medições dinâmicas não lineares. Subsidiariamente consideramos as evidências da qualificação do mecanismo dinâmico de formação da identidade de gênero seguir um padrão de criticalidade e de auto-organização – pela presença de distribuição 1/f quando comparamos intensidade e incidência no campo do gênero.

Prefácio

Quando estive em Chicago, para apresentar meu trabalho no 20º Simpósio Bianual da HBIGDA/WPATH sobre o embasamento teórico do método por nós desenvolvido para avaliação da quantificação dinâmica dos estados de variância de gênero – metodologia que nos permite desenvolver avaliações de pessoas variantes de gênero a distância pela internet de forma precisa e segura – realmente eu me senti como se eu tivesse chegado a Chicago tendo vindo não de São Paulo, mas de Marte, talvez de Vênus, mais provavelmente de Alfa-Centauro.

Minha comunicação foi difícil, não só pela língua estrangeira – com a qual estou acostumada por trabalhar quotidianamente lendo e escrevendo nela – se bem que os sotaques muito variados realmente foram para mim em alguns casos bastante indecifráveis – mas principalmente por grandes diferenças entre nossas culturas científicas – e nosso embasamento teórico.

Nós estamos muito engajados na nova ciência da não linearidade dos processos sistêmicos, onde a estratégia é sempre da complexidade, de considerar as múltiplas causas e suas interações, para estudar efeitos complexos, em sistemas que apresentam uma imprevisibilidade estrutural com a convivência de efeitos com probabilidades muito variadas – sistemas longe do equilíbrio são nossa especialidade.

Quase todos especialistas em Chicago – como no Brasil – ainda consideram como prática usual a antiga e tradicional maneira de ver um pouco reducionista, procurando causas e efeitos simples para explicar partes de fenômenos muito complexos – o que certamente cada dia mais se afasta da realidade vivida e experimentada por nós.

Por isso, verto para o português, de forma livre, uma parte do prefácio do livro de Sole e Goodwin, publicado em 2000:

"Uma formidável explosão de criatividade em ciência vai transformando as disciplinas tradicionais rapidamente, catalisando transformações de forma que antigos limites estão se dissolvendo e a integração de disciplinas vai sendo delineada. Essa nova visão ou percepção tem como base a complexidade, a situação caótica como o imprevisível e o inesperado, e a emergência de uma nova ordem. Começou na física e na matemática e agora se desenvolve rapidamente nas ciências da vida, onde está revelando novas assinaturas do processo naturalmente criativo

que embasa a evolução dos organismos. O sinal distintivo da vida é a emergência espontânea do novo a partir da complexidade material de si mesma. O conceito de emergência, no passado considerado por muitos em biologia como vago e místico através de perigosas considerações vitalistas, é agora o foco central da ciência da complexidade. Agora a questão é: Como sistemas compostos de componentes conhecidos podem dar origem a resultados e desenvolvimentos inesperados?

Cada vez mais fica evidente que nosso novo entendimento dos sistemas complexos nos leva para além da perspectiva científica tradicional da predição e do controle na natureza. Esse novo entendimento nos leva a uma relação de participação nos processos naturais, os quais mesmo sendo imprevisíveis, ainda podem permanecer inteligíveis”.

(Ricard Solé & Brian Goodwin - Signs of Life - How Complexity Pervades Biology, Basic Books, 2000)

Eu espero que essas palavras expliquem melhor o que nós estamos desenvolvendo – e o que já desenvolvemos.

Afinal, o que parece mais simples do que a diferenciação de sexo e gênero em uma dicotomia absolutamente certa e absolutamente simples?

A própria lei considera evidente essa repartição simplista com base na morfologia do que temos entre as pernas. Porém, como avaliam Solé e Goodwin, mesmo nesse caso aparentemente tão simples, que social, ideológica e mesmo juridicamente chega a ser estabelecido “a priori” como certo – surge o desenvolvimento inesperado – Joãozinho se sente e se percebe como Maria, ou Mariazinha como José... ou como um alguém ainda mais inesperado.

O objetivo principal de nosso trabalho – e de nossa apresentação no simpósio de Chicago - **não foi** o estudo das variâncias de gênero – GV nem da formação da identidade de gênero – GI – do ponto de vista de **sua etiologia**.

Ao contrário, o principal objetivo de nosso método foi desenvolvido visando a **medição**, a **quantificação** da dinâmica de formação da identidade de gênero – de forma a permitir um método de sua comparação com estados típicos – sem se considerar em nenhum momento qualquer teoria de etiologia, que teria levado a esse estado – muitas vezes inesperado.

Desenvolvemos um método de avaliação – de **quantificação objetiva** – para auxiliar a avaliação presencial – ou mesmo para permitir avaliações a distância via internet – para que se possa saber objetivamente se alguém vive um estado transexual quando pode precisar efetivamente de uma transição, tratamento

hormonal e de uma redesignação sexual (cirurgias SRS) e um novo assentamento civil; ou se alguém vive um estado transgênero (travesti) quando pode precisar de ajuda para que se promovam transições parciais e mesmo cirurgias e alguma hormonização dirigida e limitada sem cirurgia de redesignação; ou se alguém vive um estado de transformismo (cross-dressing) ou mesmo de intergênero – precisando de acompanhamento e orientação.

Nosso método não é limitado por diferentes embasamentos teóricos com relação a causas. As causas que regem o desenvolvimento da identidade de gênero sabemos serem complexas e que podem levar a estados totalmente inesperados de baixa probabilidade.

O nosso objetivo – essa quantificação – principalmente através da internet – a olhos mais desavisados pode parecer uma fantasia – como uma coisa irrealizável. Para outros pode parecer que estamos invadindo terreno alheio em psiquiatria ou psicologia – fazendo “testes psicométricos” infundados pela internet. Outros já imaginaram que estaríamos fazendo “tele-psiquiatria”. Outros imaginam avaliar nossos métodos, considerando testes de validação da psicomетria clássica – que veremos não se aplicam.

Nossa apresentação em Chicago foi em sessão plenária, mas sob o tema da tele-psiquiatria. Na realidade, foi grande a dificuldade que tiveram para perceber que o que fazemos não é tele-psiquiatria. Agora ainda resta um percurso para que entendam que o que fazemos não é tele-avaliação-psicométrica.

Nos foi proposto em 2006 – e aceitamos imediatamente – um programa piloto conjunto com uma equipe européia – para avaliarmos conjuntamente o método desenvolvido por nós – e infelizmente esse programa não se desenvolveu, não por nossa vontade.

Avaliamos que um programa sério, com uma entidade brasileira de pesquisa – uma universidade ou hospital público ou privado – que desenvolva avaliações de variâncias de gênero – seria muito bem-vindo. Apenas essa instituição além de parceira na pesquisa teria também que ser financiadora do programa – pois não podemos arcar com os custos de um programa desses.

Introdução

Pretendemos mostrar aqui as principais diferenças no embasamento matemático e lógico entre os métodos clássicos em psicometria e os métodos dinâmicos de mensuração de fenômenos complexos não lineares.

Os métodos clássicos em psicometria consideram como ponto de princípio, sistemas estáticos. Comparações de características entre um indivíduo dentro de um grupo. Define-se uma população de “normais” ou “gaussianos”, e compara-se os dados obtidos experimentalmente para um indivíduo com essa população gaussiana, através de comparações estatísticas simples de médias, variâncias, etc..

Procede-se então à validação do teste (apresentamos ao final, a cópia do que publica a wikipedia.org, em inglês, sobre testes clássicos em psicometria, apenas para referência para o leitor).

Em testes psicométricos mais sofisticados, como o IRT e os modelos de Rasch, por exemplo; o embasamento é sempre o mesmo, apenas as estatísticas consideradas são mais sofisticadas.

Esses métodos sempre consideram pontos de vista estáticos e não dinâmicos – e quando consideram os dinâmicos, as relações são lineares e não complexas.

Tradicionalmente, em psicometria, ao se avaliar séries temporais, sempre se usa métodos lineares – métodos considerados de “curta-duração”. Sempre nesses métodos existem os pressupostos subjacentes da distribuição “gaussiana” de uma “normalidade”.

Certamente esses métodos são absolutamente corretos e extremamente úteis – para o que se destinam, tomando por base seu embasamento lógico, ideológico, e seu ponto de vista linear e gaussiano. Se aplicam a inúmeros testes e são extremamente úteis não só em psicologia, mas em economia e em outras áreas das ciências.

Realmente, na natureza, quando se vive um estado estacionário, uma grande maioria de fenômenos se comporta de forma gaussiana. Mas um desenvolvimento não é uma situação de um estado estacionário – pelo contrário, o que o caracteriza é a realidade transiente, muitas vezes longe do equilíbrio – onde essa metodologia tradicional certamente não é adequada.

Aqui, não estamos tratando de estados estacionários e muito menos de populações gaussianas.

Nesse sentido, de forma alguma estamos desenvolvendo o que se costuma caracterizar como “testes psicométricos”.

O que queremos medir?

- Queremos **medir a auto-percepção de gênero**: a percepção de ser homem, mulher, ambos, nenhum dos dois, entre ambos ou alguém diferente.
 - Na realidade, dada à complexidade do sistema e a dificuldade da mensuração, nos propomos medir não a auto-percepção – **mas a dinâmica de como a auto-percepção se desenvolve no tempo**, de forma a podermos quantificar esse desenvolvimento através de séries temporais obtidas experimentalmente e posteriormente avaliadas levando-se em conta a complexidade do sistema.
 - Como um **objetivo subsidiário**, mas certamente significativo do ponto de vista teórico, procuramos também desenvolver instrumentos e obter dados **de intensidade da variância de gênero** para podermos comparar com dados publicados da incidência de estados típicos. Cremos que essa comparação é bastante importante para no futuro poder-se **compreender melhor os mecanismos envolvidos entre as variáveis** importantes no desenvolvimento da formação da identidade de gênero.
-
- Procuramos atingir nossos objetivos sabendo:
 1. Que essa percepção da identidade de gênero é função de um conjunto extenso e complexo de variáveis conhecidas e provavelmente de outras ainda desconhecidas;
 2. Ser o desenvolvimento e a estabilização (ou não) da identidade de gênero certamente imprevisível “a priori”. Evidência absoluta dessa afirmação é a incidência sistemática, em todas as culturas, da existência de variâncias de gênero de intensidade variável, desde casos extremos como de transexuais, como casos mais amenizados como travestis, bem mais amenizados como transformistas, além de casos de intergênero – a partir de situações de desenvolvimento sexual típico ou atípico (intersexo) ;
 3. Que um sistema sensível a condições iniciais, podendo gerar como resultado estados imprevisíveis com probabilidades heterogêneas constitui um sistema caótico – que pode demonstrar assinaturas

- dinâmicas reconhecíveis, legíveis e mensuráveis em seus múltiplos estados típicos.
4. Que um sistema caótico só pode ser avaliado e analisado – reconhecido, lido e/ou medido, considerando-se uma metodologia não linear (Spratt 2003);

- Por outro lado, para tanto, podemos considerar:
 1. Que não conhecemos sequer todas as variáveis desse sistema. Conhecemos algumas certamente importantes... genes, hormônios, tecidos e grupos de núcleos cerebrais basais em sua diferenciação sexual certamente são um fator muito importante. Sabemos também que o estresse continuado da mãe durante a gestação pode afetar a parte hormonal do feto de forma bastante importante durante a fase de diferenciação do cérebro basal. Por outro lado também fatores adquiridos como sexo de criação, influência cultural e familiar, interesses sexuais e afetivos podem vir a ser importantes. Certamente um sistema onde Natureza e Ambiente colaboram e interferem em aspectos muito variados, provavelmente o meio interferindo num embasamento natural bastante complexo mas precocemente estabelecido;
 2. Que certamente nesse sistema complexo, muitas causas e muitas interações entre causas geram muitos efeitos – entre eles o desenvolvimento da identidade de gênero.
 3. Que, nesse sistema complexo de estados possíveis, onde muitas causas produzem muitos efeitos – se considerarmos topologicamente esse espaço um espaço discreto e não um contínuo, em condições apropriadas de “embedding” poderemos reduzir o sistema a uma relação simplificada mais inteligível de muitas causas gerando uma variável observável – que no nosso caso chamamos de GI;
 4. Podemos/devemos considerar, dada a imprevisibilidade estrutural desse sistema, os seus estados menos prováveis e mais distantes do equilíbrio como “patologias” que fugiriam dos estados mais prováveis (nesse caso mais “normais”)? Faz sentido, num sistema intrinsecamente “a priori” imprevisível, considerar-se todo estado longe do equilíbrio – longe dos estados mais prováveis – uma situação “a priori” de patologia? Mesmo hoje em dia, quando sabemos que as situações longe do equilíbrio são essenciais para a emergência do novo?
 5. Que a imprevisibilidade natural desse sistema, dada a natural existência de estados de maior e menor probabilidade, pode evidenciar a expressão natural de auto-organização do sistema revelando uma estrutura fractal em sua variabilidade e no desenvolvimento de diversidade de estados, não necessariamente patológicos mas simplesmente diversos?

6. Que espaços ou campos, em situação de criticalidade (veja-se os exemplos teóricos do Ising e da criticalidade em sistemas que se auto-organizam – Sole & Goodwin 2000) experimentam em comparações de incidência e intensidade uma situação “fractal” expressa por uma situação conhecida como “espectro” $1/f$ (sabemos que o tema pode ser controverso em psicologia principalmente – nas outras ciências o assunto se mostra bem mais claro – vide diferentes pontos de vista na controvérsia entre Gilden 2001; Van Orsen et al (2003, 2005); Thornton & Gilden 2005; Wagenmakers et al (2004, 2005) – na realidade uma luta entre o antigo método linear de análise contra o novo método não linear, por exemplo);
7. Que por esse motivo, consideramos como um objetivo secundário – mas importante – a verificação experimental – se não da criticalidade do sistema (que ainda não podemos afirmar de forma inequívoca – mas apenas sugerir como muito provável), pelo menos da sua fractalidade, por termos evidências de que a incidência e intensidade seguem efetivamente uma distribuição fractal $1/f$ – como mostramos dados na nossa apresentação em Chicago;
8. Que só podemos – e devemos – classificar alguém como vivendo um “estado patológico” – que discrimina, marginaliza e pré-dispõe à exclusão e ao ostracismo – profissional, emocional e afetivo – além de social, quando podemos mostrar evidências inquestionáveis desse estado – devido a dados indiscutíveis de desordenamento mental. Não bastam, para uma adequada classificação de um estado “patológico”, alevisias referentes a pretensas “normalidades” estatísticas em estados de maior probabilidade quando se sabe que naturalmente o sistema pode desenvolver a emergência de estados de menor probabilidade – dentro do universo fractal da diversidade e da variabilidade natural possível, no sistema complexo do desenvolvimento de diferentes estados de identidade de gênero.

"Não podemos mais nos posicionar de forma ingênua como meros observadores externos dos fenômenos que queremos manipular. As propriedades dos organismos, sua saúde e desenvolvimento, a atividade dinâmica de cérebros e comunidades, a ordem que caracteriza os sistemas ecológicos, os padrões de mudança evolutiva, são processos nos quais estamos envolvidos. Para o melhor ou para o pior, nós participamos deles, e desejamos participar de forma sábia, e não irresponsável" (Sole & Goodwin, 2000) - minha versão livre para o português.

O objetivo principal e o subsidiário

Nosso objetivo principal, já afirmamos, foi o de desenvolver um método de medida das assinaturas dinâmicas do desenvolvimento da identidade de gênero – e não a consideração de causas e etiologia.

Pelo contrário, procuramos desenvolver um método que nos permita medir quantificando efeitos de forma segura e precisa – independentemente das causas e de sua complexidade.

Sabemos, por outro lado, que existe uma gradação nos efeitos. Não existem apenas homens e mulheres – existem também, uma gradação de masculinidade e feminilidade, que se manifestam de muitas formas típicas que são classificadas nas SOC6 da HBGDA/WPATH, que representam situações de variância de gênero numa escala de gradações de diferentes estados bem definidos e outros menos definidos – mas todos de incidência reduzida – uma baixa probabilidade – considerando-se os estados mais prováveis.

Sabemos que o sistema é complexo e que estudos simplísticos de uma causa – um efeito – simplesmente causam mais confusão do que esclarecem esse sistema.

Medidas de incidência e intensidade – nosso objetivo secundário.

Como se pode medir incidência e intensidade estudando-se o espaço de gênero?

Quanto à incidência, Conway (2002, 2007) está desenvolvendo medidas e avaliações – e tem seus dados publicados. Consideramos esses dados em nossas avaliações – na falta de melhores. Estudos mais detalhados, esperamos, venham a ser desenvolvidos no futuro.

E quanto à avaliação de intensidade?

Desenvolvemos há já alguns anos – mais precisamente em 2001 e 2002 – questionários com esse objetivo – MF9 para MtF's e FM1 para FtM's. Podemos considerar também outras escalas, como o Cogiatl – entre outras escalas possíveis, para esse fim. Também aqui, novas escalas, e o uso intensivo delas em estudos de intensidade e para avaliação do espectro, podem e devem ser desenvolvidos.

Na época nós validamos essas escalas – mas como não tinham nenhum objetivo clínico ou de avaliação – não publicamos.

Nosso maior objetivo foi o de verificar se estávamos num sistema fractal ou não, considerando-se o espaço dos diferentes estados possíveis de desenvolvimento da identidade de gênero. Verificamos – como mostramos em Chicago em nossa apresentação – que o sistema é fractal – evidenciando um espectro de intensidade e incidência do tipo $1/f$.

Evidentemente a confirmação desse resultado – e estudos mais elaborados e exaustivos certamente serão necessários para comprovar e afirmar essa evidência – muito importante para o estudo da formação da identidade de gênero em sua etiologia.

Desenvolvendo testes dinâmicos para reconhecimento de situações típicas através de assinaturas – nosso objetivo principal

Nosso objetivo principal foi o de desenvolvermos um método capaz de medir assinaturas dinâmicas típicas da formação da identidade de gênero – de forma a que possamos reconhecendo e comparando assinaturas, estabelecer de forma objetiva um método que auxilie o diagnóstico diferencial – a diferenciação dinâmica - dos diferentes estados de variância de gênero.

Medida a assinatura dinâmica do cliente – não necessariamente um paciente – através da comparação de sua assinatura com as assinaturas típicas estabelecidas para as categorias de estados previstos nas SOC 6 por exemplo – podemos dar um passo importante no estabelecimento de uma diferenciação entre o cliente e os estados possíveis.

Para entendermos as características dinâmicas de um desenvolvimento, precisamos estudar seu movimento.

O movimento, como estabelecido por Newton, está relacionado a posições e velocidades – ou gradientes. Sempre diz respeito a tempo – e espaço.

Dessa forma definimos como espaço de gênero, o espaço virtual onde se desenvolvem as identidades de gênero. Na realidade dividimos esse espaço em dois, com base no sexo de designação dos indivíduos: Espaço MtF para os designados originalmente como masculinos e FtM os designados originalmente como femininos.

No espaço de gênero – nos dois semi-espacos, a identidade de gênero se desenvolve no tempo.

Movimentos muito simples, em espaços com poucos graus de liberdade, conforme as leis de Newton, poderiam ser estudados como um contínuo, usando-se equações diferenciais. Henri Poincaré porém, no final do século XIX verificou que, quando o número de graus de liberdade nesse sistema crescem, as soluções se tornam impossíveis – os sistemas podem se tornar simplesmente intrinsecamente incalculáveis.

No início se pensava que nos faltava conhecimento e instrumentos para o cálculo. Poincaré verificou porém que o que faltava não era habilidade – mas o cálculo era efetivamente impossível – pois com o aumento do número de variáveis surgiam incertezas estruturais no sistema.

Portanto, as leis de Newton só poderiam ser calculadas quando aplicadas a sistemas simples ou muito particulares – quase todos os sistemas mais complexos eram insolúveis.

Com o desenvolvimento dos computadores e do cálculo numérico – usando-se não mais equações diferenciais em espaços contínuos mas equações de diferenças em espaços discretos e iterativos – desenvolveu-se o que se conhece hoje como a teoria do Caos determinístico, e mais recentemente desenvolveu-se o conhecimento dos sistemas longe do equilíbrio, os conceitos de auto-organização, de emergência do novo a partir de flutuações longe do equilíbrio, e a teoria dos sistemas complexos.

Calcular – e mesmo verificar topológica e graficamente – em espaços discretos e iterativos – se mostra a melhor solução – na realidade a única solução possível para a grande maioria dos sistemas.

Assim podemos estudar “mapas” – a relação do sistema com seu passado – iterativamente usando-se computadores. Computadores são hoje essenciais para o conhecimento e o desenvolvimento do conhecimento nas ciências que estudam sistemas complexos.

Podemos assim definir também um “espaço de fase” – onde podemos reconhecer o que é estocástico e o que é determinístico; o que é simples e o que é caótico, o que novo e o que é estranho.

Através de mapas e de diagramas de fase, podemos estabelecer assinaturas típicas – e estudar as características de estados típicos num espaço – o espaço de gênero por exemplo. Simplicidade, periodicidade, caos determinístico e a emergência de novidade podem ser então estudados – suas assinaturas conhecidas – e reconhecíveis – mensuráveis. Estados de desordem estrutural e de estocasticidade também podem ser facilmente reconhecidos – por sua falta de um padrão de assinaturas dinâmicas em todos os níveis.

No espaço de gênero – como em nenhum espaço – não existem “normais”, muito menos “gaussianos”. Existem estados mais ou menos prováveis, e suas assinaturas dinâmicas reconhecíveis e mensuráveis.

A validação de nossa avaliação é intrínseca ao próprio método. Comparamos assinaturas de indivíduos e famílias. Assim se faz em outras áreas da ciência e da tecnologia (Szücs et al. 2003; Amon & Lefranc, 2004 – consideram uma metodologia similar em outros ramos do conhecimento, como em neurobiologia e em física, respectivamente).

Para a quantificação de assinaturas – em indivíduos e em famílias de indivíduos, consideramos o método de quantificação proposto por Sabelli (2005), usado em

inúmeras áreas do conhecimento, como em biologia, em astrofísica, em geologia, em psiquiatria e em ciências sociais, por exemplo.

Sabemos que temos muito a conhecer, e um longo caminho a percorrer para que compreendamos de forma plena a imprevisibilidade do desenvolvimento da identidade de gênero e as variâncias de gênero. Seus mecanismos, em suas variáveis, em sua etiologia, em suas nuances e possíveis estados – mas os primeiros passos foram dados – de forma objetiva permitindo o uso de uma tecnologia nova – para a avaliação de casos de variâncias de gênero – a distância.

Classical Test Theory (Wikipedia.org)

True and error scores

Classical test theory is based on the decomposition of observed scores into true and error scores. The theory views the observed score x of person i , denoted as x_i , as a realization of a random variable X . The person is characterized by a probability distribution over the possible realizations of this random variable. This distribution is called a "propensity distribution". Person i 's true score, t_i , is axiomatically defined as the expectation of this propensity distribution. This definition is formally stated as

$$\text{(Eq. 1)} \quad \varepsilon(X_i) = t_i.$$

Secondly, the so-called [error](#) score for person i , E_i , is defined as the difference between i 's observed score and his true score:

$$\text{(Eq. 2)} \quad E_i = X_i - t_i.$$

Note that X_i and E_i are random variables, but t_i is a constant. Also note that it directly follows from these definitions that the error score has expectation zero:

$$\text{(Eq. 3)} \quad \varepsilon(E_i) = \varepsilon(X_i - t_i) = \varepsilon(X_i) - \varepsilon(t_i) = t_i - t_i = 0.$$

Relation to population

The above equations represent the assumptions that classical test theory makes at the level of the individual person. However, the theory is never used to analyze individual test scores; rather, the focus of the theory is on properties of test scores relative to populations of persons. Hence, the next step is to introduce a population-sampling scheme into the structure of classical test theory. When we assume that people are randomly sampled from a population, the true score becomes a random variable too, so that we get the (in)famous equation

$$\text{(Eq. 4)} \quad X = T + E$$

Classical test theory is concerned with the relations between the three variables X , T , and E in the population. These relations are used to say something about the quality of test scores. In this regard, the most important concept is that of *reliability*. The reliability of the observed test scores X , which is denoted as ρ_{XT}^2 , is defined as the ratio of true score variance σ_T^2 to the observed score variance σ_X^2 :

$$(Eq. 5) \quad \rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}.$$

Because the variance of the observed scores can be shown to equal the sum of the variance of true scores and the variance of error scores, this is equivalent to

$$(Eq. 6) \quad \rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}.$$

This equation, which formulates a signal-to-noise ratio, has intuitive appeal: The reliability of test scores becomes higher as the proportion of error variance in the test scores becomes lower and vice versa. The reliability is equal to the proportion of the variance in the test scores that we could explain if we knew the true scores. The square root of the reliability is the correlation between true and observed scores.

Reliability

Note that reliability is not, as is often suggested in textbooks, a fixed property of tests, but a property of test scores that is relative to a particular population. This is because test scores will not be equally reliable in every population. For instance, as is the case for any correlation, the reliability of test scores will be lowered by restriction of range. Thus, IQ-test scores that are highly reliable in the general population will be less reliable in a population of college students. Also note that test scores are perfectly unreliable for any given individual i , because, as has been noted above, the true score is a constant at the level of the individual, which implies it has zero variance, so that the ratio of true score variance to observed score variance, and hence reliability, is zero. The reason for this is that, in the classical test theory model, all observed variability in i 's scores is random error by definition (see Eq. 2). Classical test theory is relevant only at the level of populations, not at the level of individuals.

Reliability cannot be estimated directly since that would require one to observe the true scores, which according to classical test theory is impossible. However, estimates of reliability can be obtained by various means. One way of estimating reliability is by constructing a so-called *parallel test*. A parallel test is a test that has the property that, for every individual, it yields the same true score and the same observed score variance as the original test. If we have parallel tests x and x' , then this means that

$$(Eq. 7) \quad \varepsilon(X_i) = \varepsilon(X'_i)$$

and

$$(Eq. 8) \quad \sigma_{E_i}^2 = \sigma_{E'_i}^2.$$

Under these assumptions, it follows that the correlation between parallel test scores equals reliability (see Lord & Novick, 1968, Ch. 2, for a proof).

$$(Eq. 9) \quad \rho_{XX'} = \frac{\sigma_{XX'}}{\sigma_X \sigma_{X'}} = \frac{\sigma_T^2}{\sigma_X^2} = \rho_{XT}^2.$$

The estimation of reliability by the use of parallel tests is cumbersome, because parallel tests are very hard to come by. In practice the method is rarely used. Instead, researchers use a measure of internal consistency known as Cronbach's α . Consider a test consisting of k items u_j , $j = 1, \dots, j, \dots, k$. The total test score is defined as the sum of the individual item scores, so that for individual i

$$(Eq. 10) \quad X_i = \sum_{j=1}^k U_{ij}.$$

Then [Cronbach's alpha](#) equals

$$(Eq. 11) \quad \alpha = \frac{k}{k-1} \frac{\sum_{j=1}^k \sigma_{U_i}^2}{\sigma_X^2}.$$

Cronbach's α can be shown to provide a lower bound for reliability under rather mild assumptions. Thus, the reliability of test scores in a population is always higher than the value of Cronbach's α in that population. Thus, this method is empirically feasible and, as a result, it is very popular among researchers.

As has been noted above, the entire exercise of classical test theory is done to arrive at a suitable definition of reliability. Reliability is supposed to say something about the general quality of the test scores in question. The general idea is that, the higher reliability is, the better. Classical test theory does not say how high reliability is supposed to be. In the literature a value over .80 appears to be deemed 'acceptable'; a value over .90 is 'good'. Values between .70 and .80 are seen as mediocre but still defensible; values below .70 are bad. ^[citation needed] It must be noted that these 'criteria' are not based on reasonable arguments but the result of convention. Whether they make any sense or not is unclear.

Alternatives

Classical test theory is by far the most influential theory of test scores in the social sciences. In psychometrics, the theory has been superseded by the more sophisticated models in [Item Response Theory](#) (IRT). IRT models, however, are catching on very slowly in mainstream research. One of the main problems causing this is the lack of widely available, user-friendly software; also, IRT is not included in standard statistical

packages like SPSS, whereas these packages routinely provide estimates of Cronbach's α . As long as this problem is not solved, classical test theory will probably remain the theory of choice for many researchers.

See

<http://en.wikipedia.org>

for details about psychometry.

Bibliografia

Amon, A & Lefranc, M --- Topological signature of deterministic chaos in short nonstationary signals from an optical parametric oscillator --- *Physical Review Letters*, vol92, number 9, 2004;

Delignières, D; Fortes, M; Ninot, G --- The Fractal Dynamics of Self-Esteem and Physical Self --- *Nonlinear Dynamics in Psychology and Life Sciences*, 8, 479 – 510, 2004;

Gilden, D. L. --- Cognitive emissions of 1/f noise. --- *Psychological Review*, 108, 33-56, 2001;

Szücs, A; Pinto, RD; Rabinovich, MI; Abarbanel, HDI; Selverston, AI --- Synaptic Modulation of the Interspike Interval Signatures of Bursting Pyloric Neurons --- *Journal of Neurophysiology*, 89 : 1363 – 1377, 2003;

Thornton, TL & Gilden, DL --- Provenance of correlations in psychological data --- *Psychonomic Bulletin & Review*, 12 (3), 409-441, 2005;

Torres, W --- Web-evaluation of Gender Variances: background and methods --- Published as a free PDF download at www.gendercare.com, 2007;

Van Orden, G. C., Holden, J. G., & Turvey, M. T. --- Self-organization of cognitive performance.--- *Journal of Experimental Psychology: General*, 132, 331-350, 2003;

Van Orden, GC; Holden, JG; Turvey, MT --- Human Cognition and 1/f Scaling --- *Journal of Experimental Psychology : General*, 134, 117 – 123, 2005;

Wagenmakers, EJ; Farrell, S; Ratcliff, R --- Human Cognition and a Pile of Sand: A Discussion on Serial Correlations and Self-Organized Criticality --- *Journal of Experimental Psychology : General*, 134, 108 – 116, 2005;

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. --- Estimation and interpretation of 1/f noise in human cognition.--- *Psychonomic Bulletin & Review*, 11, 579-615, 2004;