FLUTUAÇÃO DE CRITÉRIOS NA AVALIAÇÃO DE REDAÇÕES *

Sérgio Costa Ribeiro **
Djalma Pessoa ***
Ruben Klein ***
Carlos Eduardo Falcão Uchôa ****
Nilma Santos Fontanive *****

I. INTRODUÇÃO

A fidedignidade das notas atribuídas às questões abertas e, basicamente, às provas de redação tem sido preocupação constante dos especialistas em avaliação e medidas educacionais.

De um lado, os instrumentos de avaliação, – quando considerados quanto ao seu valor intrínseco –, são objeto de aceitação da maioria dos educadores, que vêem neles o potencial de medir uma gama variada de habilidades intelectuais. De outro, persistem as dificuldades quanto à homogeneidade de critérios de julgamento e, assim, quanto à confiabilidade dos resultados da aplicação de tais instrumentos.

Marelim Vianna[1], em um excelente trabalho de revisão de estudos empíricos realizados sobre a fidedignidade e validade das provas de redação utilizadas como medida da capacidade de expressão escrita, atesta a pobreza de pesquisas brasileiras sobre esta área de investigação. Neste artigo, Vianna reporta-se a inúmeros estudos empíricos realizados no exterior sobre a fidedignidade deste tipo de prova, os quais analisam a variabilidade de avaliação dos julgadores, enfatizando ainda os problemas de variabilidade de desempenho de um mesmo julgador em função de diferentes momentos de julgamento de uma mesma redação.

* Trabalho parcialmente financiado pelo projeto "Vestibular: Instrumento de Diagnóstico do Sistema Escolar" FINEP cont. nº B/40/79/148/00/00.

** PUC/RJ e CESGRANRIO, *** Instituto de Matemática Pura e Aplicada - CNPq, **** Universidade Federal Fluminense, ***** UFRJ e CESGRANRIO.

(1) Marelim Vianna, H. Redação e medida da expressão escrita: algumas contribuições da pesquisa educacional. *Cadernos de Pesquisa*, São Paulo, 16: 41-7, 1976.

Em um estudo com uma amostra pequena de 161 sujeitos, Vianna[1] confirma algumas das conclusões das pesquisas internacionais anteriormente revisadas por ele.

O presente trabalho pretende ser uma contribuição ao pequeno acervo de pesquisas realizadas neste campo no Brasil. Seus autores, embora cientes da limitação deste estudo, crêem que a inclusão de provas ou questões de redação, como um dos componentes dos concursos vestibulares nos últimos 3 anos, merece exaustivas investigações no tocante aos aspectos de flutuação dos critérios de julgamento, já que tais variações podem obviamente diminuir o seu valor como instrumento eficaz de seleção, e a sua confiabilidade para discriminar candidatos.

Em particular, o experimento realizado pela Fundação Cesgranrio – que lida com vestibular de grandes números – pode permitir uma série de inferências qualitativas que justificam estudos quantitativos com os dados disponíveis nesta instituição.

II. O CONTEXTO DA FUNDAÇÃO CESGRANRIO

A possível inclusão obrigatória da redação no Concurso Vestibular de há muito preocupava a Fundação CESGRANRIO. Por esta razão, ela providenciou, em 1975, uma pesquisa sobre a viabilidade da redação no Vestibular classificatório e, com a finalidade de promover um estudo sério e eficaz sobre o problema, constituiu uma comissão formada por dez especialistas, sendo cinco de Língua Portuguesa e cinco de Medidas Educacionais. Esta comissão iniciou os trabalhos em setembro de 1975 e no dia 11 de dezembro do mesmo ano apresentou parecer técnico sobre o assunto, indicando, entre outras providências, a necessidade da realização de uma experiência brasileira, feita com grandes números, que pudesse simular, da melhor forma possível, a situação real de um exame de acesso ao ensino superior.

Uma Comissão Especial de Professores de Língua Portuguesa planejou e supervisionou um Concurso de Redação, realizado em outubro de 1976, com a participação de 10.000 candidatos.

Relembrem-se dois pontos importantes enfatizados na análise dos resultados obtidos naquela experiência: a) a expressiva diversidade de avaliação das redações, apesar da homogeneidade na amostragem dos lotes de provas distribuídos a cada dupla de professores e apesar do treinamento a que foram submetidos os 20 docentes escolhidos para a correção; b) a também expressiva correlação entre o desempenho dos candidatos no Concurso de Redação e o desempenho dos mesmos candidatos nas questões de múltipla escolha de Português do Vestibular de 77.

O Decreto nº 79.298, de 24 de fevereiro de 1977, determinou a "inclusão obrigatória de prova ou questão de redação em língua portuguesa" no Concurso Vestibular de 78.

Com base no Concurso de Redação, em estudos feitos com dados do Vestibular de 77 e em experiências realizadas com turmas de 2º grau, foram adotadas, após cuidadosa análise, as seguintes normas gerais: 1º) necessidade de alterar a escala de avaliação adotada no Concurso de Redação (de 0 a 100), a fim de se tentar alcançar uma convergência maior no julgamento da capacidade de expressão escrita dos candidatos; 2º) constatada expressiva correlação de desempenho na redação e nas questões de múltipla escolha de Português, a conveniência de se avaliar a redação através de percentuais a serem acrescidos ao resultado obtido nas questões de múltipla escolha de Português.

Estabeleceu-se então a seguinte orientação quanto ao valor da redação no Vestibular de 1978: o escore bruto, ou seja, a nota real do candidato na disciplina Língua Portuguesa e Literatura Brasileira da Prova de Comunicação e Expressão seria acrescido de 30% ou 15% de acordo com o conceito A ou B, respectivamente, obtido na redação, não tendo acréscimo o escore bruto do candidato a cuja redação fosse atribuído o conceito C.

A mesma orientação geral presidiu os trabalhos do Vestibular de 1979.

2.1. – Organização

No Vestibular de 1979, os trabalhos foram desenvolvidos com a seguinte estrutura organizacional, análoga à do vestibular de 1978: uma coordenação geral integrada por um coordena-

(1) Marelim Vianna, H. Flutuações de julgamento em provas de redação. *Cadernos de Pesquisa*, São Paulo, 19: 5-9, 1976.

dor e dois subcoordenadores; doze equipes de avaliadores com 13 membros cada uma, totalizando 155 docentes (um avaliador de uma das equipes deixou de participar da avaliação). Cada equipe, com a supervisão de um professor, trabalhando oito horas diárias durante seis dias, julgou 8.200 redações. Assim, um avaliador corrigiu em média 630 redações nestes seis dias.

Os 12 supervisores, professores de nível universitário, com longa experiência e competência comprovada, foram escolhidos intencionalmente de áreas distintas (Língua Portuguesa, Lingüística Geral e Literatura Brasileira e Portuguesa), com a finalidade de reunir docentes que possivelmente reagiriam de maneira diferente na apreciação de problemas de expressão escrita. O que se queria era provocar debates, confrontar posições, para se tentar chegar à possível unidade na diversidade.

Já os 155 professores, a quem caberia a responsabilidade de avaliar as redações, foram recrutados, na sua maioria, do ensino de 2º grau, com atuação em escolas oficiais e particulares, situadas em diferentes áreas sócio-culturais. Aqui também se visava ao encontro de vivências bem distintas. Era muito importante o conhecimento mais abrangente possível da realidade sobre a qual se iria trabalhar.

A partir de agosto de 1978, objetivando maior convergência na apreciação dos diversos problemas de uma redação, começaram as sessões de treinamento. A coordenação se reunia com os 12 supervisores e estes com as suas respectivas equipes. Para as sessões de treinamento conseguiu-se um número expressivo de redações de alunos de 2º grau, de escolas oficiais e particulares. A coordenação procedeu a uma seleção destas redações, separando para um treinamento aquelas que julgou serem, por fatores diversos, as mais problemáticas de avaliação. Pode-se dizer que, às vésperas do Vestibular, tinha-se alcançado uma convergência muito razoável quanto à atribuição dos conceitos A, B e C.

2.2 – Período de Avaliação das Redações

A redação foi aplicada no mesmo dia da prova de múltipla escolha de Comunicação e Expressão, tendo a duração de uma hora.

O local escolhido para o trabalho de avaliação foi tranqüilo, confortável e operacional. As 12 equipes trabalhavam num mesmo andar, em salas vizinhas, permitindo assim que os supervisores mantivessem contato entre si e com a coordenação.

O avaliador recebeu, por dia, 3 pacotes de 40 provas no máximo. Em dois dias (2º e 4º do trabalho), recebeu 4 pacotes. Na verdade, um pacote tinha em média 35 redações, por causa dos candidatos faltosos. Assim, era de 105 a 110 o número de redações avaliadas por dia por um professor (nos dois dias de 4 pacotes, este número foi em torno de 140 redações). Tal número não foi considerado excessivo pela quase unanimidade dos professores. Em relação a muitas provas, era fácil atribuir um dos três conceitos logo após uma primeira leitura das mesmas. As que não apresentavam maiores problemas de avaliação eram julgadas por um único professor. Havendo hesitação na atribuição do conceito, o professor solicitava ao seu supervisor que lesse a redação. Em certos casos, a própria coordenação era consultada.

A coordenação estabeleceu dois turnos de trabalho. No primeiro (das 8 às 12,30 horas), o avaliador só poderia corrigir 2 pacotes. Mesmo que terminasse o 2º pacote antes das 12,30 horas – o que foi raríssimo – só recebia o 3º pacote às 14 horas, depois do intervalo do almoço. Entre dois pacotes, o avaliador era obrigado também a um pequeno descanso.

2.3 – Critérios Gerais de Avaliação

Desde o Vestibular de 78 se chegara ao consenso (coordenação, supervisores e professores) de que na avaliação das redações se deveria dar maior ênfase à capacidade de o candidato organizar o pensamento, em relação ao tema proposto, de maneira clara e coerente. É importante enfatizar que o domínio da norma culta e do sistema ortográfico vigente não estaria de maneira alguma deixando de ser avaliado. O que sempre se defendeu, desde o início, foi dar prioridade à capacidade de organização do pensamento. Por norma culta da língua, à falta de ou-

tros estudos, considerar-se-ia basicamente a que é preconizada, com pouca variação, pelas nossas gramáticas escolares, não se deixando de levar em conta recentes pesquisas sobre a norma literária brasileira contemporânea.

Dentro desta orientação geral, foi adotada a seguinte linha de trabalho: após a leitura da redação, a ela se atribuía um dos três conceitos, com base apenas na organização do pensamento: conceito A para a redação com boa organização do pensamento; B para a redação com razoável organização do pensamento – um ou outro problema de falta de concatenação sintática, repetições desnecessárias, falta de seqüência lógica entre as idéias de um parágrafo e outro, pontuação reveladora, em uma ou outra passagem, de deficiente ordenação de idéias. . .; enfim C para a redação em que o candidato revelasse não saber pensar. A má ordenação de idéias – há, é claro, aí uma escala, da redação caótica à sofrível – penalizaria com o conceito C, independentemente da ocorrência de problemas em relação à norma culta e ao sistema gráfico. Em geral, os candidatos que mostram não saber estruturar o pensamento apresentam problemas também no tocante sobretudo ao domínio da norma culta. Quanto às redações a que foram atribuídos os conceitos A ou B, com base na organização do pensamento, tinha-se a seguir outra preocupação: manter ou não o conceito inicial dado, depois de avaliados os dois outros aspectos da expressão escrita – a norma gramatical e o sistema ortográfico. Uma redação com boa organização do pensamento devia então cair para o conceito B, caso apresentasse certo número de desvios inequivocamente graves em relação à norma culta (concordância, flexões nominais e verbais. . .) e ao sistema ortográfico (por exemplo, palavras de uso freqüente grafadas erroneamente). Do mesmo modo, uma redação com razoável organização de pensamento podia passar a merecer o conceito C se apresentasse certo número de desvios inequivocamente graves em relação à norma culta e ao sistema ortográfico vigente.

Esclareça-se, por fim, que uma redação visa a avaliar precipuamente a capacidade de uma pessoa saber expressar-se por escrito com clareza, coerência e correção e não aferir o grau de conhecimento sobre o assunto. Assim sendo, uma redação em que a organização do pensamento é boa poderia vir até a merecer conceito A, não obstante a pobreza do seu conteúdo.

III. METODOLOGIA DE ANÁLISE – TÉCNICA ESTATÍSTICA

Para a análise dos dados do experimento, foi utilizada uma técnica estatística, análise de correspondência, que é uma generalização da análise de componentes principais para dados categóricos. Seu objetivo é descrever e resumir as informações contidas nos dados através de uma redução da dimensão dos espaços considerados. Para uma descrição sucinta da técnica, consideraremos uma tabela de contingência I x J, isto é, o cruzamento de I classes A1, . . ., AI da categoria A com J classes B1, . . ., BJ da categoria B. Obtêm-se representações das classes da categoria A como pontos no espaço de dimensão J e das classes da categoria B como pontos no espaço de dimensão I. Em cada um desses espaços será gerada uma distância apropriada a fim de se poder julgar a similaridade entre as classes de uma mesma categoria. A freqüência relativa de ocorrência de cada classe é utilizada como um peso para essa classe. A seguir, aplica-se, separadamente, uma análise de componentes principais generalizada às representações das classes A., . . ., AI e às representações das classes B1, . . ., BJ. Dessa maneira, escolhida uma dimensão menor que I e J, tem-se para cada uma das análises uma representação das classes em um espaço de dimensão p de maneira a "melhor" conservar a informação dos dados. Por exemplo, se p=2, ter-se á uma representação gráfica no plano com a distância euclidiana usual. Assim, se 2 classes estão próximas na representação original, também estão próximas na representação em p variáveis e, se estão afastadas na representação em p variáveis, também estão afastadas na representação original.

Podem-se representar graficamente as classes das duas categorias no mesmo gráfico e essas duas representações são relacionadas. Uma relação existente entre elas é que a coordenada da classe Ai, por exemplo, é, a menos de um fator de expansão, uma média ponderada das coordenadas das classes Bj no mesmo eixo, e vice-versa. Logo, especialmente na periferia dos gráficos, podem-se perceber em geral quais classes Bj são mais relacionadas com quais classes Ai. Em aná-

lise de componentes principais, calculam-se também as correlações entre as variáveis originais e as novas variáveis obtidas (as coordenadas no novo sistema de referência), chamadas as componentes principais, a fim de ajudar na interpretação dessas novas variáveis. Faz-se o mesmo em análise de correspondência para cada uma das duas análises feitas. É interessante notar que as correlações das classes Bj, vistas como variáveis na primeira análise de componentes principais generalizadas, são relacionadas com as coordenadas das classes Bj na segunda análise de componentes principais generalizadas, relação esta que mantém o sinal, e vice-versa.

Para maiores detalhes sobre a técnica, assim como seu desenvolvimento matemático, remetemos o leitor à literatura estatística pertinente(1,2 e 3).

IV. ANÁLISE DE RESULTADOS

4.1 – A primeira visão geral dos resultados da avaliação das redações pode ser observada na *figura 1*, onde o número de questões acertadas pelos alunos na prova de múltipla escolha é comparado com os percentuais relativos de conceitos A, B e C atribuídos às redações. O gráfico evidencia a associação entre os escores de acertos na múltipla escolha com a atribuição dos conceitos A, B e C. A correlação de Pearson global é de 0,46.

Observa-se, por exemplo, que à medida que o número de acertos aumenta, diminui a percentagem de conceitos C e aumenta a percentagem de conceitos A. Já o conceito B é mais freqüentemente atribuído aos candidatos que acertaram cerca de 60% das questões de múltipla escolha. A média geral da prova de múltipla escolha foi de 16 acertos, isto é, 40% das questões. Os percentuais totais de A, B, C foram 12,5, 34,2 e 53,3, respectivamente.

4.2 – Utilizamos a técnica estatística, já descrita na tabela de contingência, cujas linhas são os 155 avaliadores das redações e cujas colunas são o cruzamento dos conceitos A, B e C com as 6 faixas de acertos (0-9, 10-14, 15-19, 20-24, 25-29 e 30-40) em que a prova de múltipla escolha foi dividida, que denotamos por A1, . . ., A6, B1, . . ., B6, C1, . . ., C6.

A *figura 2* mostra a representação obtida dos cruzamentos A1, . . ., C6. Os símbolos A , B , C denotam, respectivamente, as médias ponderadas dos conceitos A, B e C, enquanto as médias ponderadas das faixas de acertos 1 a 6 estão contidas no círculo hachurado com centro na origem.

A concentração em torno da origem das faixas de acertos evidencia a independência da categoria "avaliadores" da categoria "faixa de acertos" na tabela de contingência considerada. Este fato é consistente com a hipótese de que a distribuição de acertos na prova de múltipla escolha de Português é a mesma para cada avaliador.

Geralmente, quanto mais comum é a ocorrência de um cruzamento de variáveis ou de uma variável, mais próximo do centro do gráfico deverá se situar a sua representação. Isto pode ser evidenciado com os conceitos A, B e C, já que sabemos que a porcentagem desses conceitos foi na ordem descrescente de C para A. Verifica-se que os cruzamentos A6, B4 e C1 satisfazem a essa condição de proximidade do centro do gráfico e coincidem com as observações feitas a partir da *figura 1*.

Podemos interpretar o significado geral dos eixos F1 (horizontal) e F2 (vertical) da seguinte maneira: o 1º é um eixo de severidade, discriminando a atribuição dos conceitos A, B, C. O 2º eixo é também de severidade, separando mais nitidamente, no entanto, a diferença de atribuição de conceito A da atribuição de conceito B.

A técnica estatística permite agora representar cada avaliador e as médias ponderadas de cada equipe em eixos que conservam a mesma interpretação de severidade de julgamento da *figura 2*.

(1) LEBART, L. e FENELON, J. P. *Statistique et Informatique Appliquées*. Paris, Dunod, 2ª ed., 1973.

(2) BENZECRI, J. P. *L'analyse des données*. Paris, Dunod, 2ª ed., 1976.

(3) FERNANDEZ, P. J., KLEIN, R. e YOHAI, V. J. *Análise de dados multivariados*. A ser publicado.

A *figura 3* apresenta as médias ponderadas das 12 equipes, representadas pelas letras N,P,Q, R, S, T, U, V, W, X, Y e Z. Observa-se nitidamente que há uma dispersão entre as diversas equipes em relação ao critério médio que se situaria na origem das coordenadas. Nota-se, por exemplo, que a equipe N foi a mais severa, enquanto a Y, a equipe mais benevolente.

Outra inferência pertinente é que, por exemplo, a equipe Y atribuiu relativamente mais conceito A do que as outras equipes, as equipes X e R mais conceito B e a equipe N mais conceito C. Isso evidencia uma dependência entre as equipes e os conceitos A, B, C.*

Analisando agora os avaliadores em cada equipe *(figuras 4 a 15)* podem-se retirar algumas informações interessantes. Por exemplo, a equipe mais severa no julgamento (equipe N, *figura 4*) apresenta-se como razoavelmente homogênea. Homogeneidade semelhante observa-se na equipe R *(figura 7)*, embora seja esta equipe de severidade aparentemente média. Tal homogeneidade de julgamento não é observada, por exemplo, nas equipes P e W *(figuras 5 e 12)*, apesar da média ser semelhante à da equipe R.

A equipe mais benevolente (equipe Y, *figura 14*), na qual todos os seus membros são benevolentes, apresenta em especial um membro (Y1) extremamente benevolente, que associa, mais freqüentemente que seus colegas, conceito A às redações de candidatos com nota baixa na múltipla escolha. Em contraste, alguns avaliadores, como por exemplo T0 *(figura 9)*, U7 *(figura 10)*, atribuíram com mais freqüência do que seus colegas conceito C às redações de candidatos com notas altas na múltipla escolha.

Algumas especulações podem ser feitas a partir desses resultados.

Em primeiro lugar, a flutuação de critérios de avaliação pelas diversas equipes pode sugerir que o supervisor, ao transmitir à sua equipe orientação da coordenação geral, durante o treinamento e durante a correção, o faça com um grau de subjetividade que explicaria tal flutuação.

A variação de homogeneidade entre as equipes poderia estar ligada à personalidade e à atitude do supervisor em relação à sua equipe. Isso, de certa forma, foi em alguns casos constatado subjetivamente pela coordenação geral.

O grau de confiabilidade constatado no processo de correção sugere, por exemplo, que diante de uma excelente ou péssima redação, à luz dos critérios de correção previamente estabelecidos, a atribuição dos conceitos A e C, respectivamente, seria uniforme em todos os avaliadores.

Infelizmente esses não são os casos mais freqüentes. Uma redação que, à luz dos critérios estabelecidos, pudesse ser considerada média, se fosse corrigida, por exemplo, pela equipe Y (ver *figura 3*) receberia com maior probabilidade o conceito A, enquanto que a mesma redação, se corrigida pela equipe Z, teria maior probabilidade de receber o conceito B e, ainda, se corrigida pela N, talvez recebesse o conceito C. O mesmo fato poderia ocorrer dentro de uma equipe, por exemplo, na equipe W, se esta redação fosse corrigida pelos avaliadores W7, W4 e WD *(figura 12)*.

Essa não é, entretanto, toda a dificuldade. A mesma técnica estatística foi aplicada em cada um dos 6 dias de correção. É importante observar que em todos os casos a interpretação dos dois primeiros eixos foi a mesma.

O grau de severidade/benevolência representado pela ordem do eixo F1 é mostrado na *figura 16*. Testes foram feitos quanto à aleatoriedade da distribuição de redações pelas diversas equipes por dia de correção.

Observa-se neste gráfico que houve significativa variação de critérios ao longo dos dias de correção. Por exemplo, a equipe cujos dados globais indicam como a mais benevolente (Y) só o foi a partir do 2º dia. As equipes X e Z mantiveram aproximadamente o mesmo grau de severidade/benevolência a partir do 2º dia.

A equipe R, por exemplo, teve uma amplitude de flutuação grande durante os 6 dias de correção. É possível que a homogeneidade de critérios constatada na análise global desta equipe seja conseqüência apenas da média dessa flutuação.

O fato mais grave, no entanto, foi o que ocorreu na equipe V, que no 1º dia foi a 2ª mais benevolente e ao longo dos dias foi se tornando monotonicamente mais severa, até que no último dia tornou-se a 2ª equipe em grau de severidade.

---

* Uma outra análise feita, utilizando o modelo log-linear, confirmou essa asserção.

V. CONCLUSÕES

A análise qualitativa desenvolvida neste trabalho confirma os resultados de diversos estudos nacionais e internacionais a respeito da confiabilidade de julgamento e de atribuição de conceitos a questões abertas.

Utilizando uma imagem, poder-se-ia especular que esse processo de avaliação de redações seria equivalente à correção de um teste de múltipla escolha em relação à qual fossem usados diversos gabaritos, que flutuariam ainda durante o processo de computação.

Uma observação final se impõe. O critério utilizado para o controle das flutuações de julgamento foi um critério externo, ou seja, a prova de múltipla escolha. Um controle mais comum neste tipo de estudo é a utilização de médias de atribuição de conceitos de vários avaliadores independentes a uma amostra de redações.

O trabalho de Myers et alii[1] utilizou este controle, em um estudo, com uma amostra da mesma ordem de grandeza (80 mil) que a da Cesgranrio.

O quadro abaixo[2] indica que a confiabilidade cresce com o número de avaliadores. É pertinente a observação de que a confiabilidade média para um único avaliador – o caso do presente trabalho – é semelhante à correlação de Pearson medida pelos autores deste estudo nos dois anos de aplicação da prova de redação.

Este fato não é contraditório, portanto, com a hipótese de que o padrão múltipla escolha é confiável.

Fidedignidades para um e vários avaliadores

Dia	Fidedignidades*			
	Um avaliador	Dois avaliadores	Três avaliadores	Quatro avaliadores
1	0,466	0,635	0,723	0,777
2	0,364	0,533	0,631	0,695
3	0,493	0,660	0,744	0,795
4	0,476	0,644	0,731	0,784
5	0,264	0,417	0,518	0,589
Leitura total	0,406	0,577	0,672	0,732

* Fidedignidades utilizando a fórmula de Spearman-Brown.

A Fundação Cesgranrio, consciente desde o início da baixa confiabilidade da atribuição de conceitos às redações e ciente, ainda, da correlação existente entre o desempenho dos candidatos na prova de múltipla escolha de Português e o conceito que eles recebem na redação, adotou o procedimento de valorar esses conceitos a partir de percentis sobre o desempenho na prova de múltipla escolha.

Este procedimento atenua a aleatoriedade introduzida na classificação dos candidatos pela flutuação de critérios de correção das redações.

VI. AGRADECIMENTOS

Os autores agradecem a Elena Judith Ganon Garayalde pelo excelente trabalho de computação.

(1) MYERS, A.E. *et alii.* Simplex structure in grading of essay test. In: *Educational and psychological mensurement*, vol. I, 26, nº 1, 1966.

(2) *Id, ibid*, p. 45.

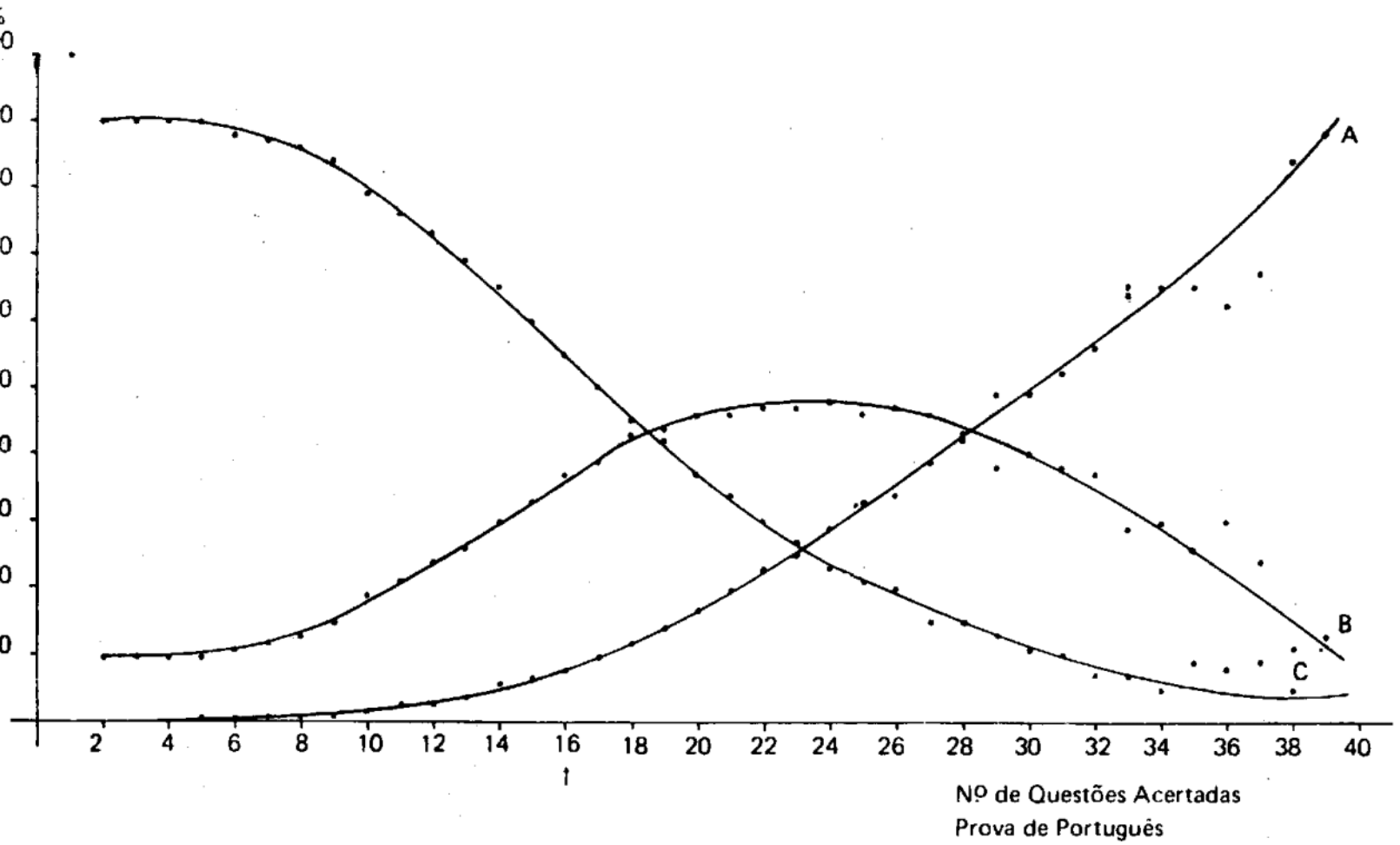
%
100
90
80
70
60
50
40
30
20
10
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
A
B
C
Nº de Questões Acertadas
Prova de Português

FIG. 1

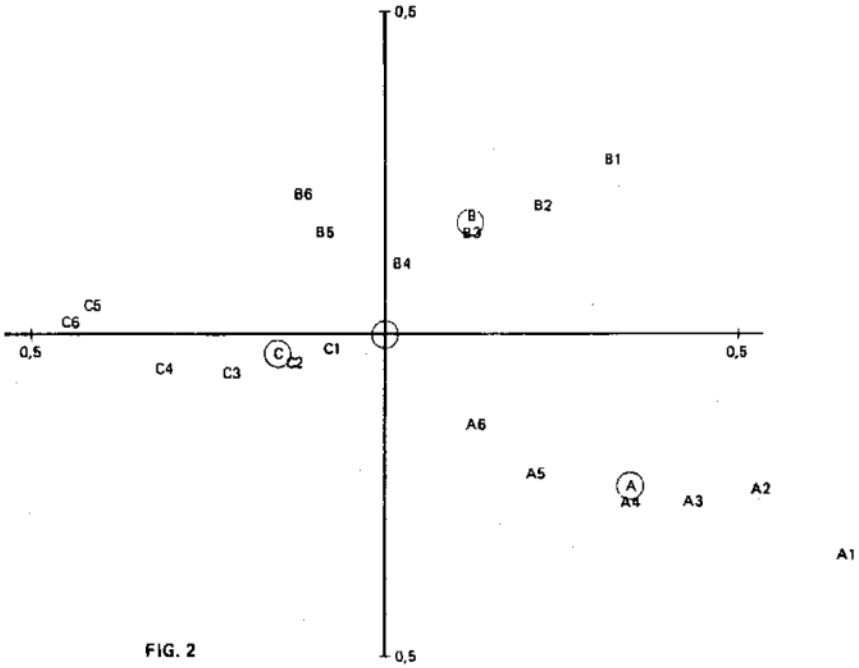
0,5
B1
B6
B2
B
B5
B4
C5
C6
0,5
C
C1
0,5
C4
C3
A6
A5
A
A2
A3
A1
0,5

FIG. 2

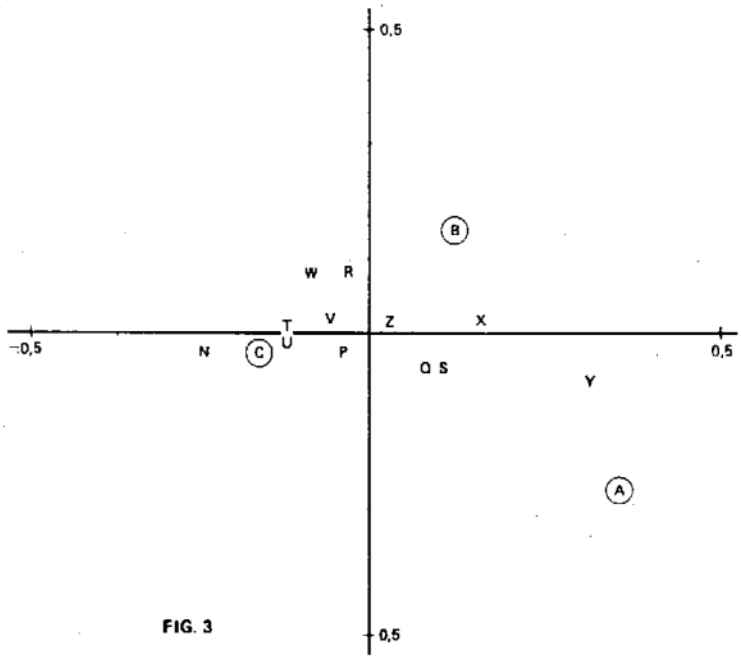
0,5
B
W
R
T
V
Z
X
−0,5
N
C
U
P
0,5
Q S
Y
A
0,5

FIG. 3

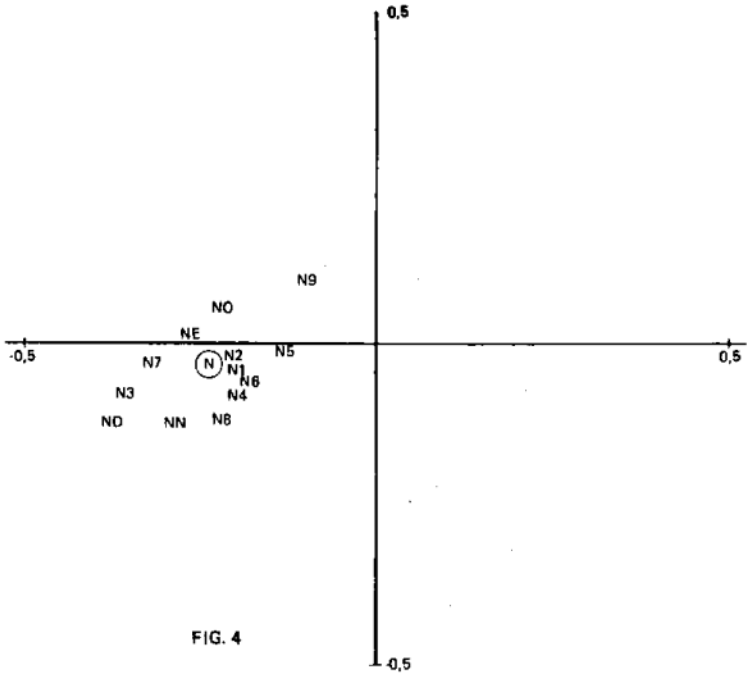
0,5
N9
NO
NE
-0,5
N7
N
N2
N1
N6
N5
0,5
N3
N4
NO
NN
N8
-0,5

FIG. 4

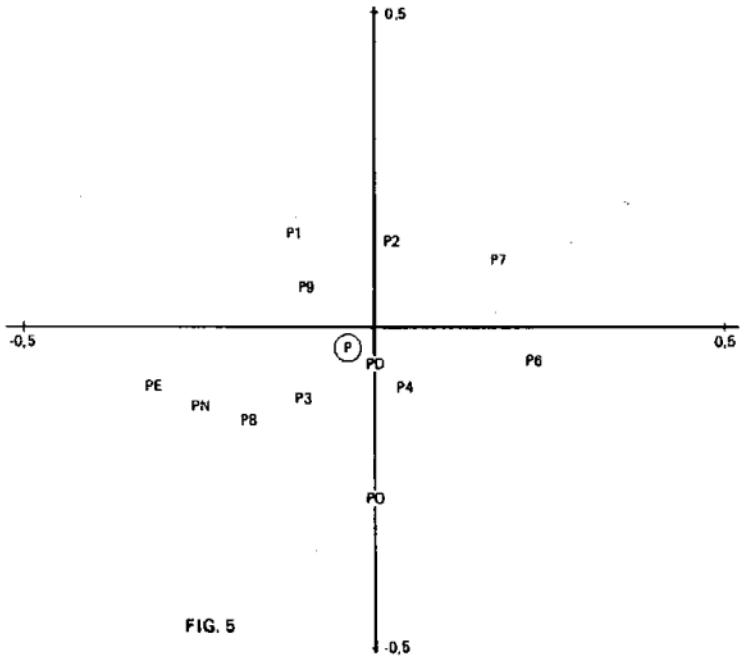
0,5
P1
P2
P7
P9
-0,5
P
PO
0,5
P6
PE
P4
PN
P3
P8
PO
-0,5

FIG. 5

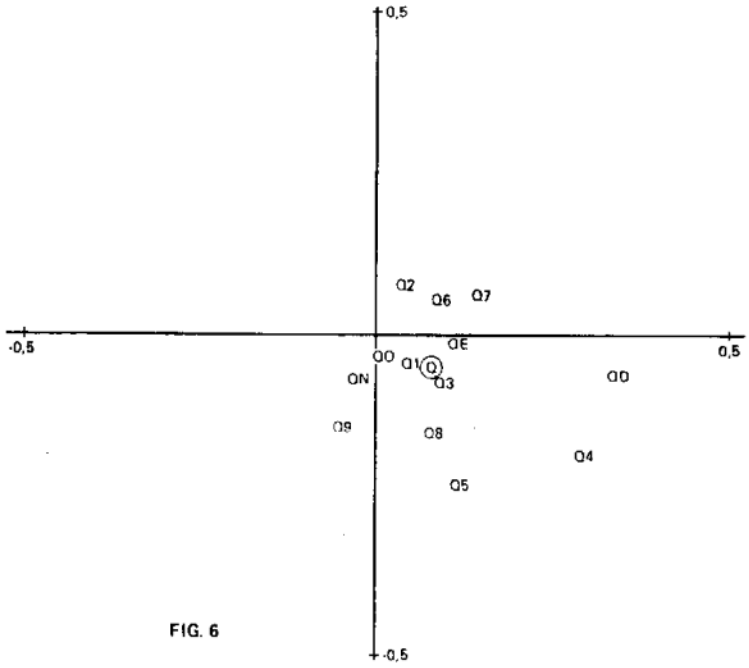
0,5
-0,5
0,5
-0,5
Q2
Q6
Q7
QE
QO
Q1
Q
Q3
QN
QO
Q9
Q8
Q4
Q5

FIG. 6

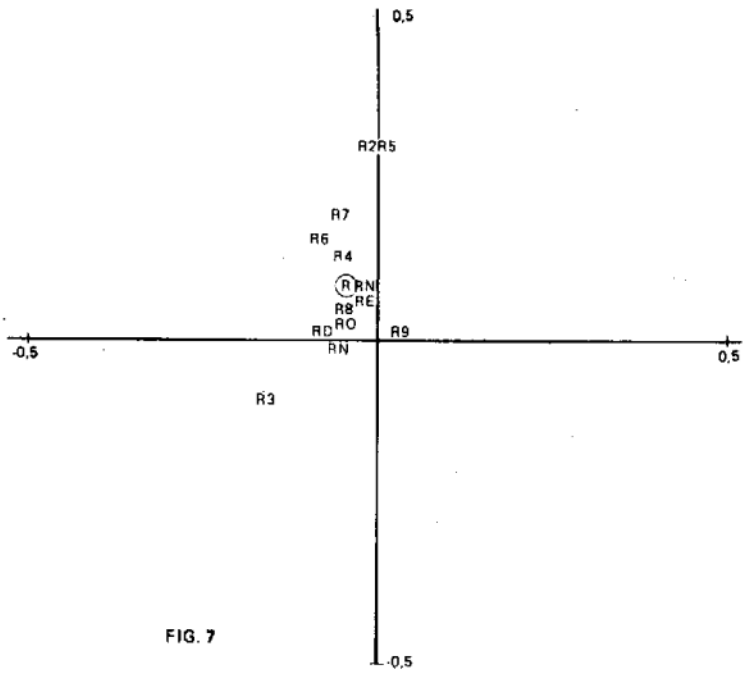
0,5
-0,5
0,5
-0,5
R2R5
R7
R6
R4
R
RN
RE
R8
RO
RD
RN
R9
R3

FIG. 7

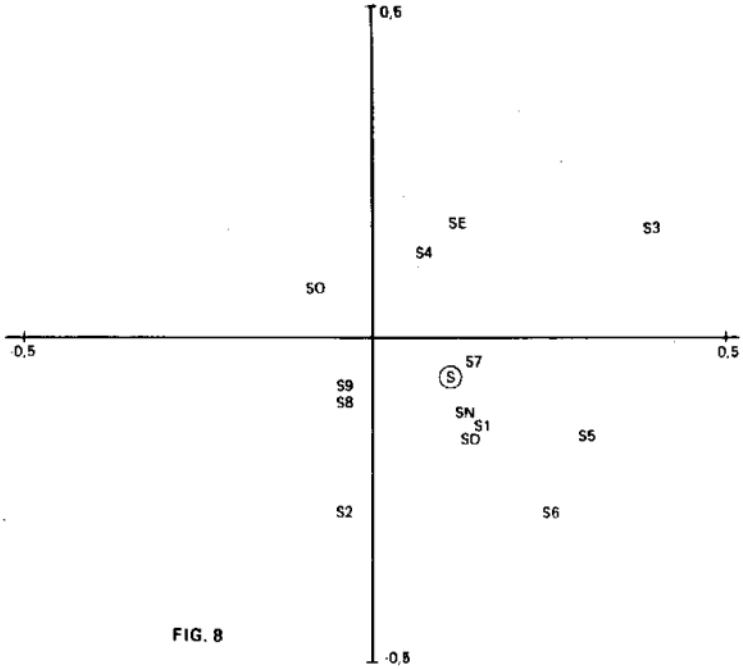
0,5
SE
S3
S4
SO
-0,5
0,5
S7
S
S9
S8
SN
S1
SO
S5
S2
S6
-0,5

FIG. 8

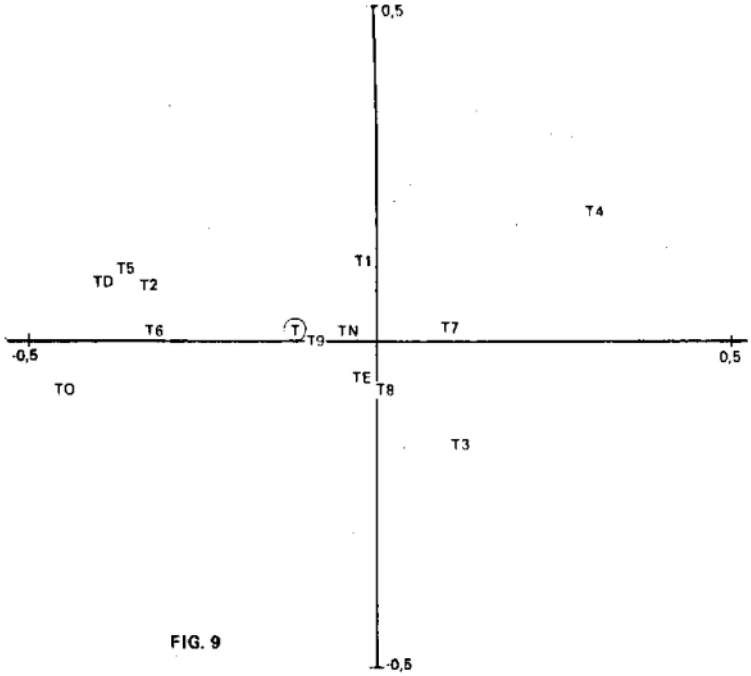
0,5
T4
T1
T5
TD
T2
T6
T
T9
TN
T7
-0,5
0,5
TE
T8
TO
T3
-0,5

FIG. 9

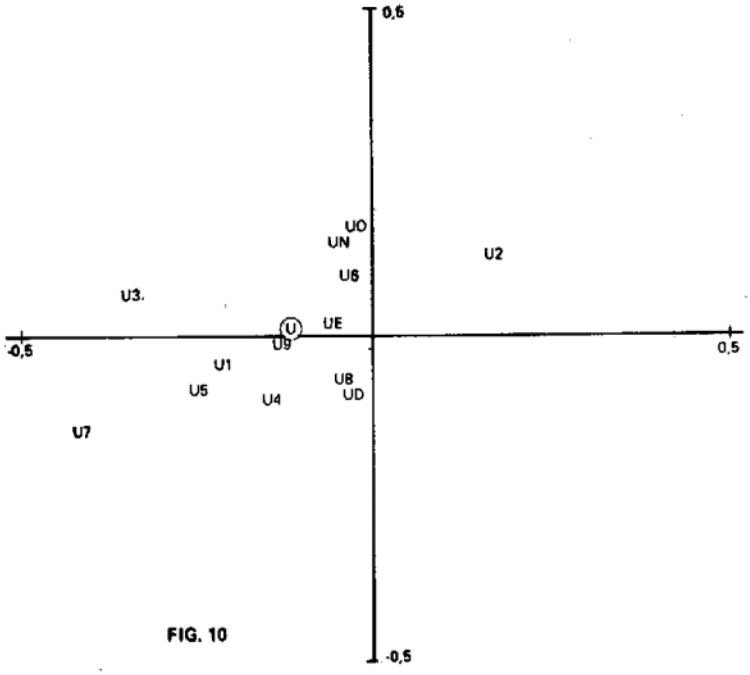
0,5
UO
UN
U2
U6
U3.
U
UE
-0,5
U9
0,5
U1
U8
U5
UD
U4
U7
-0,5

FIG. 10

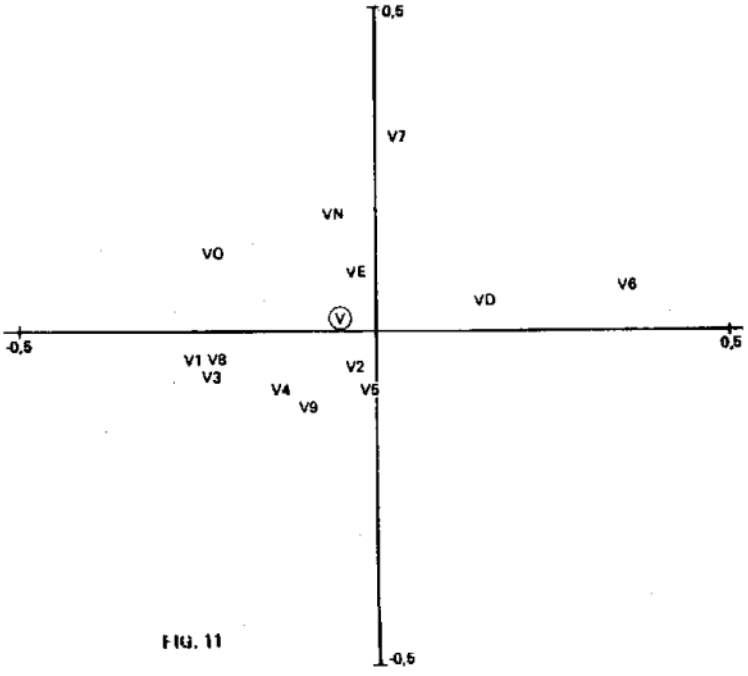
0,5
V7
VN
VO
VE
V6
VD
V
-0,5
0,5
V1 V8
V3
V2
V4
V5
V9
-0,5

FIG. 11

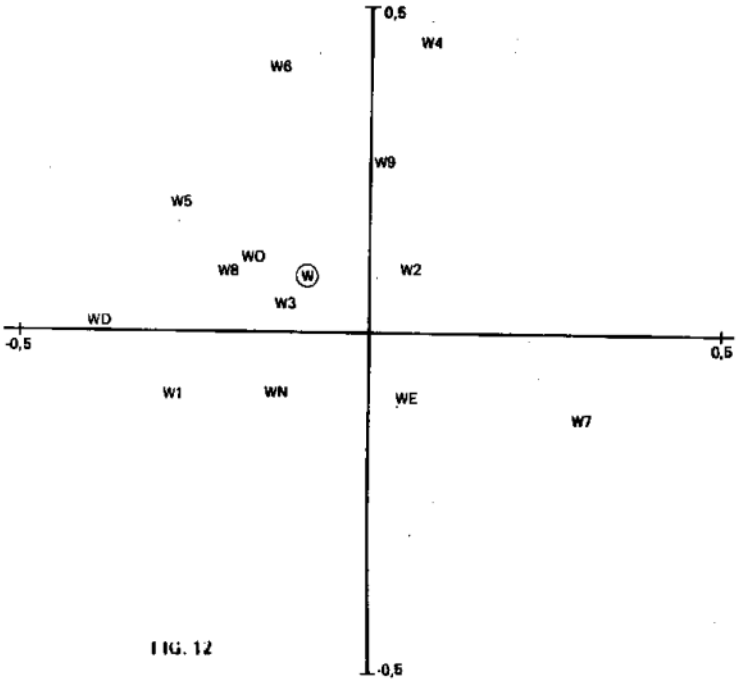
0,5
W4
W6
W9
W5
WO
W8
W
W2
W3
WD
-0,5
0,5
W1
WN
WE
W7
-0,5

FIG. 12

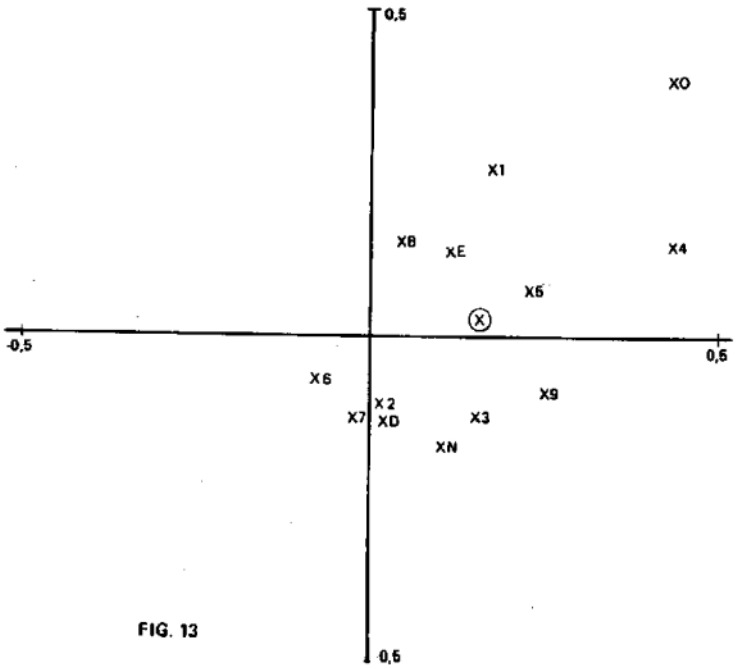
0,5
XO
X1
XB
XE
X4
X5
X
-0,5
0,6
X6
X9
X2
X7
XD
X3
XN
0,5

FIG. 13

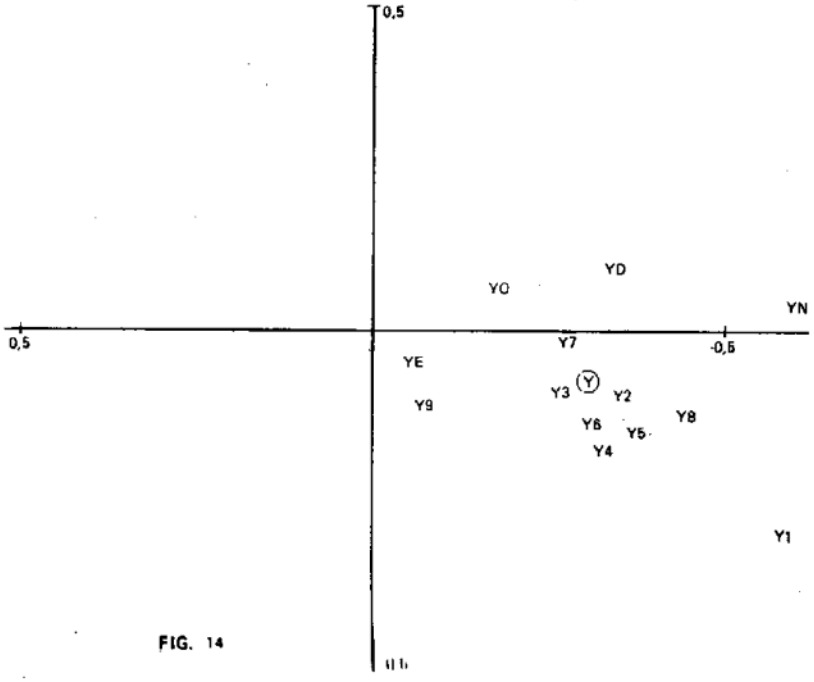
0,5
YD
YO
YN
0,5
Y7
-0,5
YE
Y3
Y
Y2
Y9
Y8
Y6
Y5
Y4
Y1

FIG. 14

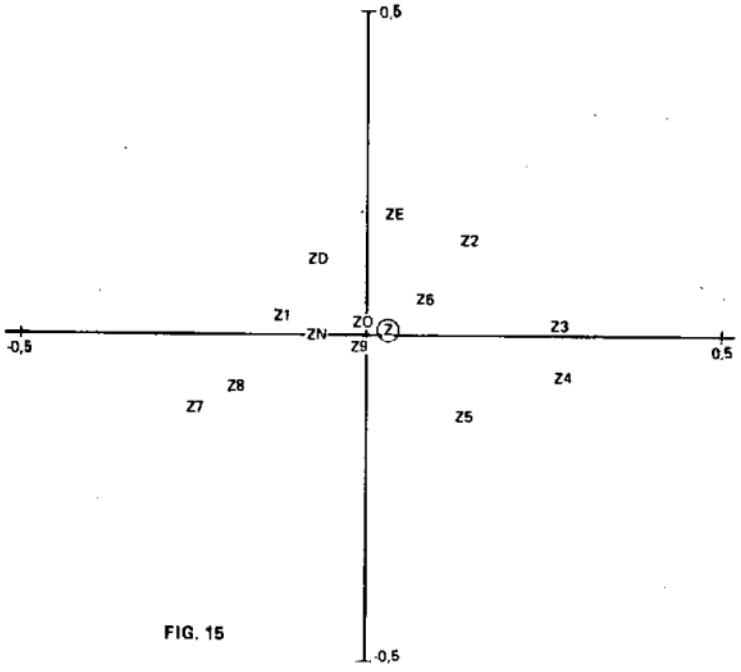
0,5
ZE
Z2
ZD
Z6
Z1
ZO
Z
Z3
-0,5
ZN
Z9
0,5
Z8
Z4
Z7
Z5
-0,5

FIG. 15

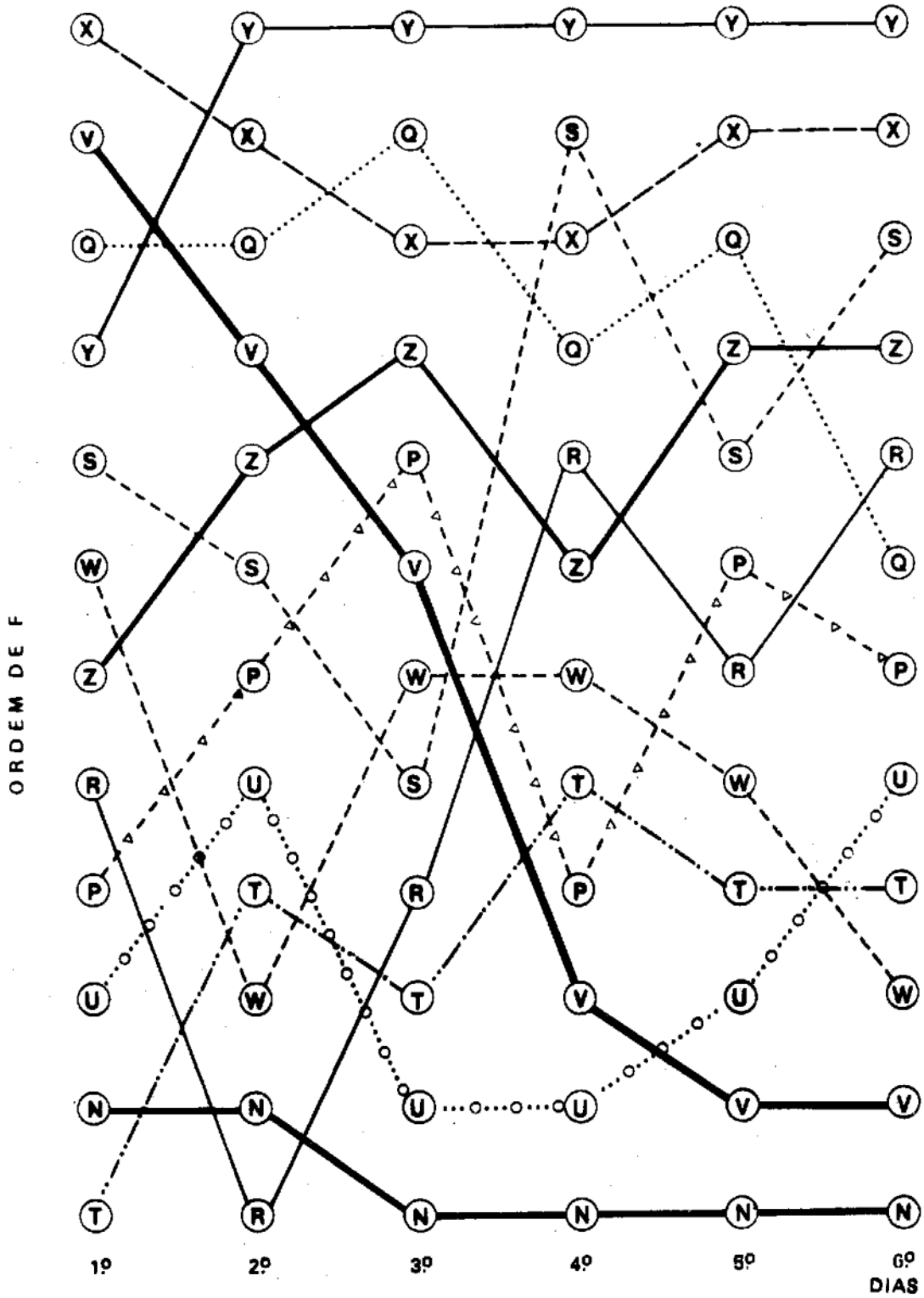

FIG. 16