

# ASPECTOS METODOLÓGICOS DO PROCESSO DE SELEÇÃO PARA O INGRESSO NAS UNIVERSIDADES BRASILEIRAS

Nícia M. Bessa(\*)

É propósito deste trabalho chamar a atenção para a necessidade de ser realizada uma avaliação objetiva do processo de seleção de alunos para o ingresso nas universidades brasileiras e, em particular, para alguns aspectos metodológicos dos exames vestibulares que não podem ser ignorados em qualquer avaliação deste tipo.

Já há alguns anos, em publicação do MEC (MEC-DAU, 1974), reconhecia-se explicitamente a necessidade de pesquisas e avaliações para fundamentar o desenvolvimento do planejamento educacional. Especificamente em relação às provas utilizadas nos vestibulares, a necessidade de informação baseada em pesquisas metodológicas ressalta dos pronunciamentos de especialistas, em simpósios, pareceres e publicações (Barroso, 1970; Bessa, 1970; Bessa 1976; Chediak, Bessa, Bechara et alii, 1975; Fernandes, 1979; Soares, 1978; Vianna, 1970). Além dos estudos realizados anualmente pela Fundação Carlos Chagas e pela Fundação Cesgranrio sobre as provas dos respectivos vestibulares e de alguns estudos metodológicos (Barroso, 1973; Buchweitz, 1976; Sanchez et alii, 1979; Vianna, 1976; Vianna, 1978), poucas têm sido as publicações de estudos empíricos que focalizam aspectos metodológicos da construção e da interpretação dos resultados das provas empregadas na seleção de alunos que demandam as universidades brasileiras.

À falta de informações objetivas, a discussão sobre os problemas envolvidos nos exames vestibulares tem sido deslocada do nível puramente técnico para o domínio da retórica e da opinião baseada na experiência pessoal. Provavelmente, terão contribuído para esse estado de coisas o incipiente desenvolvimento da avaliação de projetos educacionais no Brasil e a fragilidade da tradição de pesquisa e ensino na área das medidas educacionais.

Não é demais, portanto, discutir a finalidade e a natureza de uma avaliação objetiva dos

---

(\*) Da Pontifícia Universidade Católica do Rio de Janeiro.

exames vestibulares, assim como alguns problemas metodológicos com relação às provas utilizadas no processo de seleção. A qualidade dessas provas é importante para a eficiência do processo de seleção. As provas, por si sós, não podem responder pela eficiência do processo — que depende de um conjunto de variáveis próprias do contexto em que se realiza cada seleção. Entretanto, é importante a avaliação de sua qualidade, pois no sistema brasileiro os resultados obtidos nessas provas constituem a única base para a discriminação entre os que devem e os que não devem ingressar nas universidades.

Na primeira parte do presente trabalho, são focalizadas algumas características do processo de seleção para o ingresso de alunos nas universidades brasileiras, que são relevantes para uma avaliação do sistema. A seguir, são discutidos aspectos metodológicos dos estudos para avaliação da discriminação entre candidatos de desempenho superior e inferior nas provas dos vestibulares, dos estudos de fidedignidade e de validade das notas obtidas nessas provas.

## AVALIAÇÃO DO PROCESSO DE SELEÇÃO

A avaliação de projetos educacionais tomou grande impulso, principalmente nos Estados Unidos, a partir de 1960. O vulto dos investimentos realizados nesse país em projetos educacionais inovadores, a complexidade desses programas e o desenvolvimento da tecnologia atuaram, entre outros fatores, motivando a orientação de novas metodologias de avaliação de programas educacionais. Passou-se a ligar a avaliação ao processo decisório; ou seja, o problema básico da avaliação passou a ser a coleta das informações necessárias para que se pudesse tomar uma decisão sobre continuar ou não um programa educacional, alterá-lo ou dá-lo por terminado. Trata-se não apenas de definir objetivos, de medir comportamentos, de verificar a congruência entre objetivos visados e resultados obtidos; a coleta de informações é dirigida, também, para a finalidade de formulação de alternativas de decisão que possam fundamentar, com realismo, uma política educacional. Embora a metodologia da avaliação de programas educacionais continue suscitando discussões (Coulson, 1978; Kennedy, 1978; Pettigrew, 1978; Rogosa, 1978), uma característica predomina na aplicação de uma variedade de modelos de avaliação: a preocupação com o controle rigoroso na obtenção de informações e com a adequação da análise (Feshbach, 1978; Stufflebeam et alii, 1971, 26-32; Weisberg, 1978)<sup>1</sup>. Sem se confundir com a pesquisa educacional, da qual se diferencia por servir a fins diversos e por focalizar programas específicos, a avaliação de programas educacionais utiliza o mesmo rigor e os mesmos processos empregados nos vários tipos de pesquisa educacional.

É nesse sentido que uma avaliação dos processos utilizados nas universidades brasileiras para a admissão de alunos poderia servir de base para a tomada de decisões realistas. Não seria simplesmente uma discussão das qualidades ou deficiências das provas empregadas nas seleções operadas pelas diversas universidades; nem apenas o resultado do trabalho de comissões, cujas opiniões, embora valiosas, são expressas em pareceres inevitavelmente permeados de subjetividade. A avaliação de programas, atualmente, requer o emprego de especialistas para a formulação de modelos adequados, para o controle da coleta de informações e para a análise dos dados (Stufflebeam et alii, 1971, 302); no caso particular do acesso aos cursos superiores brasileiros, a presença de especialistas em medidas educacionais torna-se indispensável, dado o papel que as provas desempenham na seleção dos candidatos. É o rigor metodológico que assegura resultados confiáveis para que possam ser sugeridas alternativas de decisão sobre o processo de seleção de alunos para as universidades, de tal modo que as decisões sobre alterações a serem introduzidas no processo se fundamentem na verificação objetiva das vantagens e desvantagens do atual sistema.

---

(<sup>1</sup>) Atualmente, membros de várias associações americanas preparam os Standards for Evaluation of Educational Programs, Projects, and Materials a serem publicados em breve.

Quando se fizer uma avaliação objetiva dos processos de seleção utilizados para o ingresso nas universidades brasileiras, será importante verificar com têm sido solucionados, na prática, vários problemas metodológicos envolvidos na construção e na interpretação dos resultados das provas empregadas. Entretanto, essas provas são utilizadas como parte de um processo mais amplo e sua avaliação não pode ignorar certas características do conjunto (Bessa, 1976, 18-22). Duas dessas características, das mais relevantes para uma avaliação do sistema de seleção, são focalizadas a seguir.

## **PROCESSOS DE SELEÇÃO PRATICADOS POR DIFERENTES INSTITUIÇÕES**

Não há um sistema único de seleção para o ingresso em nossas universidades; há diversos sistemas, praticados por diversas instituições. Esses sistemas podem ter características comuns, principalmente em relação a práticas estabelecidas por lei ou pelo uso. Entretanto, diferem basicamente quando apresentam um conjunto diferente de provas para cada instituição, um conjunto diferente de opções para ingresso nos cursos superiores e quando o próprio conjunto de cursos oferecidos difere quanto a diversos aspectos.

Assim sendo, não teria sentido uma avaliação que não considerasse cada sistema praticado para a seleção de alunos, separadamente. Isto é, não teria sentido avaliar-se o sistema de ingresso nas universidades brasileiras em geral; isto equivaleria a pretender-se avaliar um sistema abstrato, em lugar dos casos concretos. No sentido que se lhe dá, atualmente, em educação, a avaliação de programas considera cada caso particular, coletando informações relevantes para a tomada de decisões sobre cada um deles. Não se avalia, portanto, um sistema abstrato de seleção, mas os diversos sistemas existentes, separadamente; ou seja, avalia-se o sistema de seleção da PUC-RJ, ou o da USP, ou o das universidades filiadas à Fundação Carlos Chagas, ou o daquelas filiadas à Fundação Cesgranrio, por exemplo.

De outro lado, os aspectos em relação aos quais os sistemas de seleção diferem nas várias instituições tornam-se relevantes para uma avaliação dos mesmos na medida em que podem afetar os resultados obtidos pela seleção.

Numa revisão da literatura sobre seleção de pessoal, Bray e Moses (1972) observam que ao iniciar-se a década de 70 já era crescente a tendência a considerar-se uma variedade de elementos nos modelos de seleção: considerar-se não somente as medidas de aptidão, mas desde o treinamento dos candidatos até a natureza da organização para a qual se opera a seleção. Desse ponto de vista, é importante que uma avaliação da eficiência da seleção para o ingresso em cada curso universitário, feita através de vestibulares de cada instituição, considere a formação dos candidatos, a composição da população de candidatos que se apresenta a cada ano, o sistema de opções para cursos e o sistema de vagas a preencher, assim como o currículo e o ambiente do curso oferecido pela universidade. Tanto quanto as disciplinas e a natureza dos exames incluídos no processo de seleção, esses são fatores que podem afetar o resultado obtido. Ou seja, uma avaliação do processo de seleção para determinado curso universitário não pode considerar apenas o custo do processo e as provas incluídas, mas também uma série de aspectos, inclusive características do próprio curso, os quais podem afetar a eficiência da seleção de alunos.

## **CRITÉRIOS PARA AVALIAÇÃO DA SELEÇÃO**

Cada curso universitário tem objetivos definidos, seja explicitamente, seja implicitamente na estruturação dos programas, na metodologia do ensino, nos processos de verificação do rendimento de aluno etc. O desempenho dos alunos no sentido de atingir esses objetivos constitui um dos critérios para julgar os resultados da seleção para o ingresso no curso universitário.

A avaliação tem papel importante na análise e na definição de objetivos imediatos ou últimos que os cursos universitários pretendem atingir (Scriven, 1967). No caso específico do processo de seleção para ingresso dos alunos no curso, cabe também ao avaliador tornar claras as

interrelações entre esses objetivos, o contexto social em que o curso se insere e o processo de seleção. Soluções apropriadas e aceitas por uma sociedade podem não ser adequadas nem aceitas por outras. O volume editado por Ottobre (1976), reunindo os trabalhos apresentados no congresso de 1976 da *International Association for Educational Assessment*, dá uma idéia da diversidade de processos utilizados para o acesso ao ensino superior em diferentes países e das condições sociais em que se inserem. No caso particular das universidades brasileiras, cabe ao avaliador estudar e explicitar o papel da universidade na sociedade brasileira contemporânea, as expectativas da sociedade em relação à universidade e o sistema de acesso adotado pelas universidades para seleção de alunos. É esse tipo de avaliação que poderá servir de base aos planejadores e aos executores de uma política educacional em relação aos cursos superiores, para que possam julgar alternativas de decisão e decidir com realismo.

A avaliação dos resultados da seleção para o ingresso em um curso universitário é problema a ser tratado empiricamente. Uma vez definidos os objetivos e os critérios para a seleção dos alunos, trata-se de verificar se os critérios são satisfeitos e se os objetivos são atingidos. Nem sempre resulta clara, para os educadores, a relação dos objetivos imediatos de um curso universitário com o papel que a universidade deve desempenhar na sociedade e com o sistema de seleção de alunos para nela ingressar. Entretanto, é importante que se tenha em vista essa relação para que se possa julgar os resultados obtidos na seleção de candidatos a uma universidade, assim como para julgar a propriedade dos critérios de seleção empregados.

Um exemplo bastante explorado da relação que deve existir entre os objetivos dos cursos universitários e o sistema de seleção para o ingresso de alunos é dado por Philip Du Bois (1964) em sua palestra de 1964 na conferência anual sobre testes e medidas educacionais patrocinada pelo *Educational Testing Service*. Du Bois comenta que cerca de 1000 anos antes de Cristo, na falta de instituições universitárias, os chineses possuíam um sistema de formação de elites intelectuais através da preparação e da seleção operada por uma série de exames para o ingresso no serviço público. No século XIV, esses exames tomaram a forma de uma competição que atraía milhares de candidatos, que eram avaliados quanto à habilidade de escrever em prosa e verso. Du Bois observa que os chineses, já então, reconheciam que o desempenho em condições controladas cuidadosamente permite estimar o desempenho futuro, no trabalho; além disso os exames abriam oportunidades iguais a todos os homens, havendo grande cuidado em conservar a igualdade de condições durante as provas, realizadas sob fiscalização severa. Em certas fases dos exames, a preocupação com a objetividade de julgamento levava a fazerem-se cópias dos trabalhos escritos, que eram codificados, sendo cada um submetido ao julgamento feito, independentemente, por três juizes. Este tipo de seleção, baseada no desempenho verbal, serviu bem à sociedade chinesa até o século XIX; foi então que passou a mostrar-se inadequado para contribuir para o tipo de formação requerido para aqueles que deveriam impulsionar o desenvolvimento científico e tecnológico do país no mesmo sentido do mundo ocidental. Mudaram os critérios, isto é, mudaram as tarefas que deveriam ser executadas pelos indivíduos cujo desempenho deveria ser previsto com base nos resultados dos exames de seleção. Os exames tiveram que ser mudados também.

Um outro exemplo serve para esclarecer a importância da relação entre os objetivos dos cursos universitários e o sistema de seleção para a admissão de alunos: uma análise, ainda que superficial, do sistema de ensino superior norte-americano mostra diferenças marcantes de objetivos, que o distinguem do sistema brasileiro. Ficam também evidentes as dificuldades para a adoção de um dos sistemas de seleção como modelo para o outro país. Ao descrever os objetivos dos cursos universitários americanos, ao nível do "college", Jameson e Ottobre (1976) distinguem inicialmente dois tipos de instituições: o "community college" e o tradicional "college" de três a quatro anos. Apesar da flexibilidade que caracteriza os dois tipos de instituições, o primeiro tem o objetivo principal da formação profissional, feita em dois anos, servindo às necessidades da comunidade que o sustenta. O segundo requer, do aluno, um longo período dedicado ao aperfeiçoamento de sua educação geral, conquanto já o prepare para cursos superiores profissionalizantes. Neste último caso, terminado o curso secundário, quando ingressa na universidade americana, com cerca de dezoito anos de idade, que tipo de formação

deve o aluno receber? Espera-se que ele complete sua educação geral, que faça uma escolha vocacional e, se desejar, que se prepare para ingressar em um curso de graduação — por exemplo, medicina, ou matemática, ou história etc. Os “community college”, em sua maioria, não são seletivos, aceitando quase todos os estudantes que terminam o curso secundário e desejam neles ingressar. Quanto ao “college” de três a quatro anos, ainda segundo Jameson e Ottobre, grande parte deles adota um processo de seleção que inclui desde notas no curso secundário, testes de aptidão, até entrevistas e cartas de recomendação.

Os programas oferecidos ao aluno que ingressa em universidades brasileiras têm objetivos diferentes daqueles do “college” americano. O aluno brasileiro que termina o curso secundário já deve fazer uma escolha vocacional, antes de se inscrever para a seleção pelos exames vestibulares, pois deve optar por um curso de medicina, ou de engenharia, ou de biologia etc. Este já é um ponto importante para o sucesso da seleção, pois pode refletir-se no desempenho do aluno na universidade; no sistema educacional brasileiro, o aluno que termina o curso secundário aos dezoito anos de idade deve fazer uma escolha vocacional cujos erros podem comprometer seu futuro por vários anos, quando se trata de ingressar em um curso superior. Perante a sociedade brasileira, a universidade assume a responsabilidade, dentre outras, de formar profissionais competentes. Dessa incumbência as universidades brasileiras devem cuidar desde o momento do ingresso de aluno em seus cursos. Se há de haver um processo de seleção, os objetivos do processo são diferentes daqueles do sistema americano, posto que os objetivos do “college” são diferentes dos objetivos da universidade brasileira. O sistema de seleção deverá ser diferente, portanto, em vista dos propósitos a que deve servir. O critério principal para uma avaliação dos resultados da seleção de alunos para um curso universitário é o desempenho em um curso profissionalizante — bem diferente do critério americano, em que o desempenho do aluno é verificado em relação a disciplinas requeridas para o aperfeiçoamento de uma educação geral.

Ao avaliar os resultados da seleção, no caso brasileiro, é importante considerar que cada curso universitário apresenta um conjunto específico de disciplinas, que podem ser comuns a alguns cursos dentro de uma mesma área de ensino, mas que divergem em geral, dos cursos de outras áreas. O rendimento escolar dos alunos deve ser medido em relação às disciplinas oferecidas em cada curso, para servir de critério de avaliação dos resultados da seleção. Obviamente, esse critério varia de um curso para outro. O avaliador deve considerar que as provas utilizadas nos processos de seleção devem servir para predizer o desempenho em conjuntos diferentes de disciplinas.

## **A QUALIDADE DAS PROVAS EMPREGADAS NOS VESTIBULARES**

São focalizados nesta seção alguns problemas metodológicos, que são universais e que têm sido objeto de estudos durante o últimos cinquenta anos, pelo menos. Nesse período, desenvolveram-se, na América e na Europa, uma tecnologia e uma teoria estatística dos testes educacionais e psicológicos. Em 1950, com o livro de Gulliksen, aparece a primeira obra em que é sistematizada a teoria dos testes educacionais e psicológicos — hoje denominada de “teoria clássica”. Em 1968, o livro de Lord e Novick apresenta algumas teorias estatísticas dos testes, além da sistematização da teoria clássica.

É no contexto da teoria clássica que são discutidos, aqui, alguns problemas relativos à fidelidade e à validade dos escores ou notas obtidas nas provas empregadas nos vestibulares e à discriminação que podem conseguir entre o desempenho superior ou inferior dos candidatos.

## **DISCRIMINAÇÃO E DISTRIBUIÇÃO DAS NOTAS OBTIDAS**

Houve tempo em que predominava uma certa mística, entre educadores, em relação à forma da distribuição de frequência das notas obtidas por um grupo de alunos, em uma prova. Es-

pera-se que em grandes grupos de alunos, que apresentassem certa heterogeneidade quanto ao conhecimento adquirido sobre determinado assunto, as notas obtidas em uma prova bem elaborada, versando sobre tal assunto, tivessem uma distribuição aproximadamente normal – ou seja semelhante à da Figura 1.

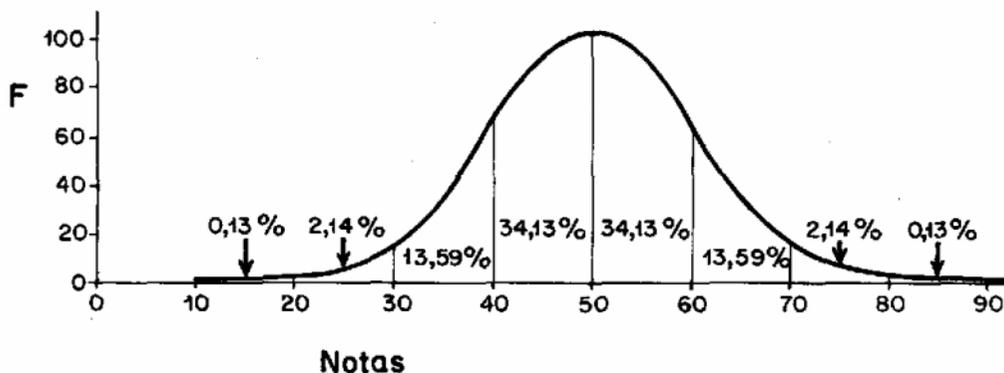


Figura 1 – Exemplo de Distribuição Normal de Frequência de Notas de uma Prova.

Não há razão para se esperar que a distribuição de notas de uma prova seja aproximadamente semelhante à Figura 1, nem para pensar que a prova não foi bem elaborada quando se obtém uma distribuição diferente. Na realidade, quando se deseja selecionar os melhores candidatos entre os que são superiores quanto ao desempenho na prova, procura-se obter uma distribuição como a da Figura 2.

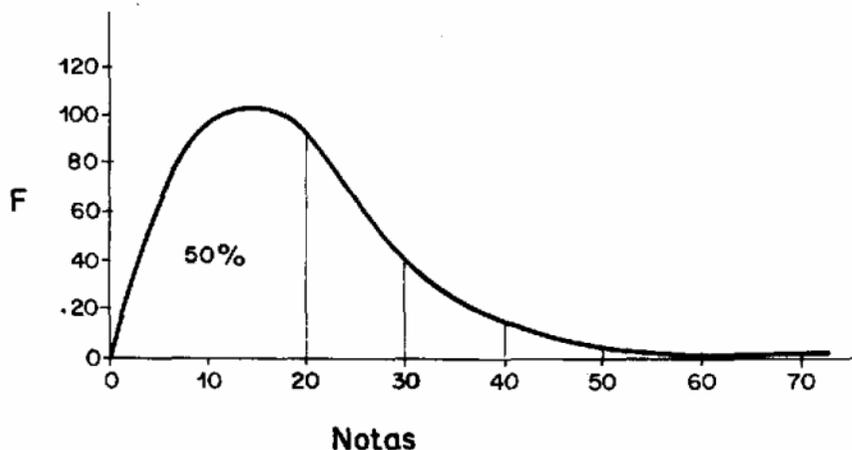


Figura 2 – Exemplo de Distribuição Assimétrica Positiva de Frequência de Notas de uma Prova.

É uma distribuição de forma semelhante à da Figura 2 que se deseja obter quando se deve selecionar, por exemplo, os 20% superiores, dentre os candidatos, com base no desempenho em determinada prova.

A forma da distribuição de frequência obtida depende do grau de dificuldade das questões e da intercorrelação entre elas. É dosando a dificuldade das questões e suas intercorrelações que se procura obter a distribuição de notas que poderá proporcionar uma discriminação satisfatória, em determinado nível de desempenho, entre os candidatos superiores e os demais (Cronbach e Warrington, 1952; Guilford, 1954, 360).

Em relação às provas utilizadas nos vestibulares, será importante verificar até que ponto as distribuições das notas dos candidatos permitem identificar os melhores, no desempenho na prova, dentro de cada grupo que escolhe determinado curso. Note-se que sendo as provas iguais para todos os cursos oferecidos, não há possibilidade de serem planejados de modo a se obter, ainda que aproximadamente, uma distribuição de forma tal que favoreça a identificação dos candidatos de desempenho superior dentre aqueles que optam por determinado curso. Na realidade, os responsáveis pela construção das provas empregadas nos vestibulares encontram-se na difícil contingência de preparar um conjunto único de provas que discriminem satisfatoriamente os candidatos de desempenho superior, dentre os inscritos em Medicina, por exemplo; ao mesmo tempo, essas provas devem identificar, igualmente bem, os candidatos superiores dentre os inscritos em Enfermagem, ou em Serviço Social, ou em Fono-Audiologia, por exemplo (Bessa, 1976). Sabendo-se que uma questão pode ser fácil para os inscritos em Medicina e bastante difícil para os candidatos inscritos em outras áreas, compreende-se que os problemas encontrados na dosagem da dificuldade das questões de cada prova do vestibular sejam difíceis de contornar.

É, pois, importante que, numa avaliação dos resultados obtidos pelo processo de seleção adotado para o ingresso nas universidades, seja estudada a forma da distribuição das notas atribuídas aos candidatos que se inscrevem em cada curso. Isto permitirá julgar se entre os candidatos inscritos em cada curso, os resultados obtidos em cada prova distinguem satisfatoriamente quais os alunos superiores, dentre os que revelam melhor desempenho. Este tipo de estudo fornecerá informações que poderão fundamentar uma avaliação objetiva da forma unificada dos vestibulares, do ponto de vista metodológico.

Evidentemente, a distribuição de notas obtidas pela aplicação de uma prova varia de um grupo de indivíduos para outro. As sucessivas avaliações dos exames vestibulares poderão indicar os reflexos de alterações nas características das populações de candidatos nos resultados dos exames. Menos óbvias, mas igualmente importantes, são alterações na própria escala de notas, que se introduzem mesmo quando as provas pretendem medir um mesmo atributo ou conhecimento.

Provas semelhantes, construídas com o objetivo de medirem os mesmos conhecimentos, podem apresentar alterações na escala de notas. De outro lado, os resultados obtidos pelos indivíduos aos quais se aplicam podem refletir alterações nas características das populações que a elas se submetem. O estudo destes dois tipos de alterações é extremamente importante numa avaliação dos resultados dos vestibulares; isto porque um exame superficial das diferenças entre grupos que se submetem, em épocas diferentes, às provas dos vestibulares pode levar a conclusões indevidas. Alguns exemplos podem esclarecer pontos básicos em que reside a dificuldade destes estudos.

A Figura 3 mostra duas distribuições de frequência de notas supostamente obtidas por dois grupos diferentes em um mesmo teste objetivo de conhecimento.

Sendo as notas dos dois grupos do exemplo da Figura 3 obtidas pela aplicação da mesma prova, não há problema em relação à comparação de médias das notas dos grupos; tampouco há problema na comparação do efeito da distribuição dos conhecimentos requeridos pela prova, nos dois grupos, sobre a seleção dos 10% de alunos superiores dentre os indivíduos que compõem cada grupo.

Quando se trata de comparar notas obtidas por duas populações diferentes em duas provas diferentes — ainda que as provas pretendam medir os mesmos conhecimentos — enfrenta-se

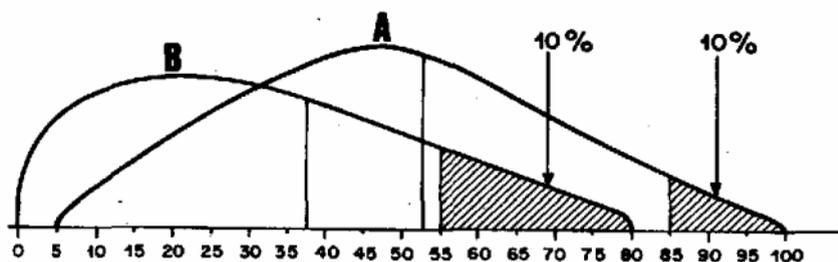


Figura 3 – Exemplo de Duas Distribuições do Escores, Obtidos Supostamente por Dois Grupos, A e B, aos quais a Mesma Prova Teria Sido Aplicada.

o problema de se ter escalas de notas que não são comparáveis. Nem sempre alterações na escala de notas parecem evidentes ao pesquisador. No caso das provas de vestibulares, que são produzidas periodicamente pela mesma instituição, pode-se ter a intenção de medir os mesmos conhecimentos e de construir provas aproximadamente equivalentes. Entretanto, um exame menos superficial poderá detectar alterações na escala de notas de uma prova a outra, quando ambas pretendem medir os mesmos conhecimentos.

Além de alterações na escala de notas, uma avaliação cuidadosa deve estudar possíveis alterações na composição das populações de candidatos que se apresentam a vestibulares sucessivos de uma mesma instituição, posto que tais alterações podem ser responsáveis por diferenças observadas nos resultados das seleções operadas. Serve de exemplo a pesquisa realizada por Beaton, Hilton e Sckrader (1977), detectando alterações na escala de escores dos testes e alterações nas características das populações examinadas e seus efeitos nas médias das notas obtidas em anos diferentes. Essa pesquisa investiga possíveis explicações para o declínio observado nos escores médios do Scholastic Aptitude Test (Educational Testing Service, 1968, 17) nos últimos anos. Embora o declínio fosse observado em várias áreas, o estudo limita-se aos escores dos testes de leitura. Para verificar diferenças, em leitura, entre alunos de escolas secundárias e entre alunos de “college” no período de 1960 a 1972, utilizaram-se dados coletados por duas pesquisas realizadas em épocas diferentes: o *Project Talent* e o *National Longitudinal Study* (Beaton et alii, 1977, 3-4). Os testes empregados nas duas investigações – a primeira, de 1960 e a segunda, de 1972 – eram diferentes e foi preciso realizar um estudo empírico para equacionar os escores de leitura obtidos nos dois estudos e torná-los comparáveis. Tornou-se possível, assim, constatar o declínio havido no período estudado, em relação às médias dos escores da leitura; a diferença confirmou-se tanto em relação às médias dos estudantes dos últimos anos da escola secundária, como em relação aos alunos do “college”. Os resultados do estudo indicam que a população do “college” não apresentava alterações em suas características, tais que sugerissem uma explicação para o declínio nas médias dos escores dos grupos em leitura. Na escola secundária, porém, os dados mostravam que a evasão de alunos era menor nos últimos anos; ou seja, alunos mais fracos, que em 1960 tenderiam a ter abandonado a escola, nela permaneciam, modificando a composição da população do secundário, em 1972. Essa tendência dos alunos mais fracos a permanecerem na escola secundária até as últimas séries, em 1972, é apontada como um dos fatores que parecem contribuir para o declínio da média dos escores de leitura entre 1960 e 1972.

De outro lado, o estudo de Beaton et alii procedeu a um ajustamento das escalas dos testes aplicados na seleção de candidatos a alguns “colleges”. O *Scholastic Aptitude Test* é requerido na seleção para o ingresso em um grande número de universidades, mas não por todas elas.

Trata-se de testes que são construídos periodicamente, pretendendo ser semelhantes, mas que não são paralelos. Foram constatadas diferenças entre as escalas de 1960 e de 1972, sendo necessário um ajustamento para torná-las comparáveis. Também na parte verbal desses testes constatou-se um declínio nas médias obtidas nos dois anos, depois de ajustadas as escalas.

Ao concluir este exemplo, note-se que o estudo de Beaton et alii ilustra dois pontos que são pouco explorados nos estudos que procuram explicar diferenças entre populações que, em épocas diversas, são submetidas a provas que são semelhantes na aparência, nos objetivos e no conteúdo, mas que não são paralelas. O primeiro ponto é a necessidade de serem examinadas possíveis diferenças entre as escalas e de torná-las comparáveis quando isto for indicado. O segundo ponto refere-se à composição das populações comparadas e seus efeitos sobre as médias obtidas nas provas a que são submetidas. Nem sempre diferenças entre médias de diversos grupos em testes de rendimento escolar podem ser explicadas por diferenças na qualidade de ensino, ou por falhas no processo de seleção. Na pesquisa de Beaton e colaboradores o declínio das médias dos testes de leitura, no "college" não pode ser atribuído a falhas na seleção, pois nem todas as universidades usam o mesmo processo de seleção de candidatos nem as mesmas provas. Uma vez ajustadas as escalas, foi possível comparar e procurar explicar a diferença entre as médias dos *Scholastic Aptitude Tests*; o estudo conclui que esse declínio reflete um menor rendimento dos alunos do secundário em leitura, estimando que a diminuição do rendimento é responsável por uma diferença de 12 pontos na média do *Scholastic Aptitude Test*, na parte verbal. Do outro lado, o declínio na média dos testes de leitura, nas últimas séries do curso secundário, em parte, se deve a diferenças observadas na composição da população escolar.

## FIDEDIGNIDADE DAS NOTAS DAS PROVAS DOS VESTIBULARES

A discussão que se vem travando, no Brasil, a respeito da preferência que se pode dar a provas discursivas, ou a provas de resposta livre, ou a provas objetivas prende-se fundamentalmente ao estudo da fidedignidade (ou precisão) das notas atribuídas aos indivíduos que se submetem a um ou a outro tipo de prova. Um breve retrospecto sobre o desenvolvimento dos estudos em relação a esses tipos de provas educacionais poderá servir para esclarecer o conceito de fidedignidade.

As críticas à subjetividade e à imprecisão das notas conferidas a provas de tipo discursivo contribuíram, em larga escala, para que surgissem as provas objetivas de conhecimentos, na América, por volta de 1920 (Chediak, Bessa, Bechara et alii, 1975; Vianna 1976). Muito antes, as questões de tipo objetivo haviam sido introduzidas em testes que tinham por objetivo medir a inteligência. Depois dos primeiros estudos na América e na Europa, o trabalho de Binet trouxe uma contribuição definitiva para o desenvolvimento das medidas de inteligência, na primeira década deste século. A escala de Binet-Simon sofreu sucessivas revisões e foi adaptada, na América, por Terman, empregando-se o conceito de quociente intelectual pela primeira vez em um teste psicológico (Anastasi, 1961, 191). Na Inglaterra também os estudos sobre medidas de inteligência tiveram grande desenvolvimento nesse período inicial do século; as contribuições de Spearman e de Cyril Burt são extremamente importantes. Já a essa época, Burt publicava resultados de "surveys" indicando uma correlação de 0,32 entre os resultados obtidos por crianças em testes de inteligência e o nível sócio-econômico dos pais; de suas investigações, conclui, ainda, que a renda auferida depende do nível de inteligência (Burt, 1962, 237-239). É interessante notar a semelhança dessas conclusões com outras, derivadas de análises de dados realizadas de modo diferente e em trabalhos bem recentes. Os estudos de Duncan, Featherman e Duncan, por exemplo, publicados em 1972, indicam que homens do mesmo nível educacional e tendo ocupações semelhantes têm renda que varia de acordo com seu nível intelectual (Duncan et alii, 1972, 102). A necessidade de aplicar testes a grandes grupos de pessoas estimulou o aparecimento de testes de lápis e papel, com perguntas de tipo objetivo, que pretendiam medir inteligência. O teste chamado Army Alpha, por exemplo, contém itens objetivos e foi aplicado largamente a grandes grupos de recursos, pelo exército americano, durante a primeira guerra mundial.

Os testes educacionais incorporaram os itens objetivos, para a medida de rendimento escolar. Relembrando os primeiros tempos da produção de testes educacionais na América, Buros (1977) descreve como o excesso de confiança nos testes objetivos e a falta de pesquisas e de crítica, levaram ao uso indiscriminado dos testes e à construção pouco cuidadosa de uma enorme quantidade deles (a relação é de cerca de 520 existentes em 1927, para cerca de 1500 atualmente). Pela década de 30, a baixa qualidade generalizada levou os testes objetivos ao descrédito. Buros considera que nessa época iniciou-se o período de maturidade, de estudos empíricos e teóricos que permitiram o desenvolvimento das medidas educacionais, principalmente na América e na Inglaterra.

Com as provas objetivas, procura-se eliminar uma das fontes de erros de medida: as diferenças entre examinadores ao conferir notas à prova de um aluno; ou as diferenças entre critérios utilizados por um mesmo examinador, em diversos momentos quando confere notas às provas dos alunos. Entretanto, a precisão dos escores de uma prova depende também das questões que a compõem, do número de questões, da composição do grupo em que as notas são obtidas, da proporção de respostas dadas ao acaso pelos alunos, das condições de aplicação etc. Não é possível eliminar todas as fontes de erros que se introduzem na medida do rendimento escolar.

Dos postulados da teoria clássica dos testes educacionais e psicológicos deduz-se a definição de fidedignidade como sendo a correlação entre escores obtidos por uma população de indivíduos em duas formas paralelas de um teste. Esta correlação é igual à proporção da variância dos escores verdadeiros no total da variância dos escores observados. Ou seja:

$$R_{XX'} = \frac{s_v^2}{s_v^2 + s_e^2} = \frac{s_v^2}{s_x^2}, \text{ onde}$$

$R_{XX'}$  = correlação entre os escores observados, obtidos por uma população em dois testes paralelos X e X';

$s_v^2$  = variância dos escores verdadeiros dos indivíduos da população, nos testes paralelos X e X';

$s_e^2$  =  $s_e^2$  = variância dos erros de medida, na distribuição de escores observados da população nos testes paralelos X e X';

$s_x^2$  =  $s_x^2$  = variância dos escores observados, obtidos pelos indivíduos da população nos testes paralelos X e X'.

A proporção da variância verdadeira na variância dos escores observados pode ser estimada de várias maneiras. Utilizando modelos de análise da variância, por exemplo, pode-se arranjar experiências controladas de modo a verificar quais os efeitos de certas fontes de erros. Este tipo de estudo é ilustrado pela pesquisa de Myers, McConville e Coffman (1966), que procura estimar a fidedignidade, ou precisão, das notas atribuídas por vários professores a provas de redação. Nessa investigação, quando se estima a fidedignidade das notas conferidas por um só professor às redações, tem-se que cerca de 53% da variância observada são devidos a erros de medida; quando é estimada a fidedignidade das médias das notas conferidas por quatro professores, as mesmas redações, essa proporção baixa para 22%.

Nos casos de uso de provas discursivas aplicadas a grande grupos de indivíduos, o custo operacional envolvido na atribuição de notas leva à utilização de um só professor para examinar e conferir nota a cada prova. Com a subjetividade desse tipo de julgamento, a variância devida a erros de medida tende a aumentar; neste caso, a fidedignidade dos escores obtidos tende a diminuir, obviamente. O estudo de Heraldo M. Vianna, com provas de redação, fornece um exemplo, entre os poucos realizados no Brasil (Vianna, 1978).

Nas provas objetivas suprime-se a subjetividade do julgamento do examinador como fonte de erro de medida, permanecendo as demais fontes, entre as quais estão as respostas assinaladas ao acaso. Note-se que o coeficiente de fidedignidade das notas obtidas pelos candidatos a

ingressar em determinado curso, em um vestibular, pode ser interpretado como um indicador da contribuição das diversas fontes de erros de medidas; nos casos das provas objetivas, entre essas fontes estão as respostas dadas ao acaso. Se fosse possível manter constantes todas as fontes de erros de medida exceto a tendência a responder ao acaso, nos testes de múltipla escolha dos vestibulares, quanto mais forte essa tendência se manifestasse entre os candidatos maior seria sua contribuição para a variância dos erros de medida; evidentemente, tenderia a baixar a fidedignidade dos escores obtidos pelos candidatos. Como exemplo, na Tabela 1, citam-se coeficientes de fidedignidade dos escores obtidos em provas objetivas empregadas nos vestibulares da Fundação Carlos Chagas.

**TABELA 1**

**Médias, Desvios-padrão, Coeficientes de Fidedignidade (Kuder-Richardson 20) e Erros-padrão da Medida dos Escores Brutos dos Testes do Vestibular da Área Biomédica da Fundação Carlos Chagas, em 1974 (N = 1667)**

Teste	Número de Itens	Média	Desvio-Padrão	KR 20	E.P.M.
Matemática	65	19.96	7.89	.81	3.45
Estudos Sociais	75	37.30	10.48	.86	3.86
Inglês	75	29.02	12.30	.90	3.78
Francês <sup>a</sup>	75	34.09	16.11	.95	3.71
Português	100	48.18	12.98	.88	4.55
Química	75	27.57	11.32	.89	3.82
Física	75	26.67	10.75	.88	3.78
Biologia	75	28.91	9.90	.85	3.83

<sup>a</sup> Baseando em 714 candidatos que fizeram a prova de Francês, não tendo sido incluídos na amostra, com os demais candidatos.

Nota - Estimativas baseadas em amostra aleatória, da população de candidatos que fizeram o vestibular da Fundação Carlos Chagas, Área Biomédica, 1974. Adaptado de Bessa, 1976. Fonte original: Vianna, 1974.

Na Tabela 1, o coeficiente de fidedignidade de 0,88 da prova de Português, pode ser interpretado como indicando que 88% da variância das notas observadas no grupo de candidatos são devidos a diferenças do conhecimento da matéria abordada pelas questões da prova; ou seja, que apenas 12% da variância dos escores observados são devidos a erros de medida de fontes diversas, entre as quais estão as respostas dadas ao acaso pelos candidatos.

É extremamente importante que se cuide da qualidade das questões, sendo este um ponto fundamental em relação à precisão das notas de uma prova. Questões mal elaboradas, tecnicamente falhas, introduzem erros nos escores obtidos. Assim também, uma prova objetiva, bem elaborada que seja, aplicada inadequadamente a um grupo de pessoas cujo nível de conhecimentos é muito inferior ao requerido pelas questões, pode apresentar resultados com alta proporção de respostas dadas ao acaso. Enfim, o controle das diversas fontes de erros de medida deve ser exercido para que as notas apresentem confiabilidade satisfatória.

Os autores da quarta edição dos *Standards for Educational and Psychological Tests*<sup>2</sup> (Davis, 1974, 50) consideram essencial a recomendação, que fazem, de que sejam fornecidos aos usuários vários tipos de estimativas, para que se possa julgar se os escores de um teste são suficientemente fidedignos para serem utilizados com o fim a que se propõem. A apresentação dessas estimativas é de especial importância quando os escores obtidos por um grupo de pessoas devem servir para fundamentar decisões que irão afetar cada indivíduo, numa seleção em que alguns são aceitos e outros são rejeitados. O problema da precisão das notas obtidas, seja em provas objetivas, seja em provas cujas notas são atribuídas por juizes, é crucial num processo de seleção, pelas implicações que tem na justiça das decisões sobre quem ingressa ou não na universidade e, até mesmo, na preservação da respeitabilidade do processo.

Os Standards (Davis, 1974, 51) fazem uma recomendação, também considerada essencial, que tem especial relevância em relação aos vestibulares realizados para o ingresso em universidades brasileiras. Essa recomendação diz respeito a testes que são comumente empregados aplicando-se a subgrupos que não são homogêneos; neste caso, as estimativas que permitem julgar a fidedignidade dos escores devem ser obtidas em relação a cada subgrupo. Justamente uma das características das populações de candidatos que têm se apresentado nos vestibulares é a diferença de nível de conhecimentos entre os subgrupos que pretendem ingressar em cada um dos diversos cursos superiores (Bessa, 1976, 12-13). Se os grupos que fazem opções por cursos diferentes, no vestibular, têm nível geral de conhecimentos diferente, é importante que sejam obtidas estimativas de fidedignidade com base nos escores dos candidatos de cada um desses grupos em cada prova.

A fidedignidade é condição necessária, embora não suficiente, para que os escores atribuídos aos indivíduos submetidos a uma prova sejam válidos. Daí a preocupação com estudos que forneçam informações a respeito da precisão das notas obtidas em cada prova que se aplica a cada população, ainda mesmo em situações em que não se trata de tomar decisões sobre indivíduos ou sobre grupos, mas de conhecer o desempenho da população no teste.

## VALIDADE DAS INTERPRETAÇÕES DAS NOTAS OBTIDAS EM UMA PROVA

As notas obtidas por um grupo de indivíduos em uma prova podem apresentar alto coeficiente de fidedignidade e, mesmo assim, podem não ser válidas as interpretações que delas são feitas. A imprecisão dos escores obtidos implica em não serem válidos; entretanto, a alta precisão não significa necessariamente que as interpretações que se fazem dos escores sejam válidas.

A pesquisa sobre a validade dos resultados obtidos por aqueles que se submetem a uma prova deve considerar vários tipos de interpretações que deles podem ser feitas. Com base no mesmo conjunto de escores, podem ser feitas interpretações diversas, algumas válidas, outras não.

Com relação às provas utilizadas nos vestibulares, é preciso ter bem claro o propósito com que são empregadas, quando se investiga a validade das interpretações das notas obtidas. Ainda que a resposta pareça óbvia, uma análise atenta da situação que se espelha nas discussões sobre os exames vestibulares revela problemas; cabe, pois, perguntar para que são utilizadas as provas que fazem parte do processo de seleção para o ingresso nas universidades brasileiras.

Se o propósito é verificar o rendimento do aluno que termina o segundo grau, interessa verificar até que ponto as provas refletem os objetivos e o conteúdo dos programas vigentes. A investigação da validade do conteúdo, ou curricular, é a mais importante neste caso. Não se trata de verificar se os objetivos de uma taxonomia, como a de Bloom (Bloom et alii, 1956) — que

---

(<sup>2</sup>) A quarta versão dos *Standards* foi elaborada por uma comissão de doze membros da *American Psychological Association*, da *American Educational Research Association* e do *National Council on Measurement in Education*.

se tornou quase obrigatória no Brasil —, estão presentes na organização dos itens das provas. Trata-se de verificar se os objetivos dos programas em uso, se o conteúdo abordado nas diversas disciplinas do currículo estão devidamente representados nas provas. O uso de uma taxonomia, como a de Bloom e colaboradores, na organização da prova, não significa que objetivos importantes do programa utilizado nas escolas, em relação à disciplina focalizada na prova, estejam presentes. A investigação da validade curricular diz respeito aos objetivos e tópicos que são realmente considerados importantes pelos professores de uma disciplina.

Não é fácil verificar quais são os objetivos visados pelos programas abordados pelos professores do segundo grau; tampouco é fácil enquadrar os itens das provas nos objetivos e tópicos de um programa de ensino. Em geral, os professores têm dificuldade em enunciar objetivos com clareza e precisão. Do outro lado, é difícil encontrar concordância perfeita entre juízes que devem classificar os itens de uma prova pelas categorias de objetivos educacionais visados. Em artigo recente, Seddon (1978) cita oito estudos que utilizam entre dois e vinte e dois juízes para classificar questões de provas segundo as categorias da Taxonomia de Bloom e colaboradores (1956); as percentagens de perfeita concordância entre juízes, obtidas nos diversos estudos, variam entre zero e noventa.

A metodologia da pesquisa da validade curricular das provas de rendimento escolar evidentemente está a requerer estudos especiais; a variedade de resultados e a divergência entre as conclusões a que chegaram diferentes estudos, que utilizam a Taxonomia de Bloom para classificar os objetivos visados pelos testes de rendimento escolar em várias disciplinas, indicam que há necessidade de novas investigações, antes de se pretender chegar a uma conclusão sobre o assunto.

Se os escores obtidos nas provas utilizadas na seleção para o ingresso nas universidades brasileiras devem ser interpretados no sentido de fazer uma previsão do rendimento dos alunos em cada curso universitário, então o exame da validade preditiva é essencial.

Quando os escores dos testes devem ser interpretados no sentido de prever o rendimento dos alunos nos vários cursos superiores, espera-se que contribuam para que a seleção resulte em um máximo de acertos. Sempre haverá erros na seleção; entretanto, espera-se que os escores dos testes contribuam para que seja mínima a proporção de falsos positivos (indivíduos aceitos e que não têm sucesso nos cursos) e de falsos negativos (indivíduos rejeitados e que teriam sucesso no curso). Para tanto, os escores dos testes empregados na seleção devem ser válidos para a previsão do rendimento no curso.

Diversos tipos de investigação empregam-se na verificação da validade preditiva dos escores dos testes utilizados na seleção de pessoal. É comum encontrar, em manuais de testes de aptidão, relações de pesquisas que verificam a correlação entre escores obtidos nos testes e medidas de rendimento escolar ou de desempenho no trabalho. Na linha de estudo de correlação entre escores de testes e notas obtidas em diversos cursos, situam-se alguns trabalhos publicados no Brasil, entre os quais estão os de Barroso (1973), Barroso et alii (1973), Bessa (1973), Bessa (1971), Bessa e Mettel (1965) e Monteiro (1964). Quando outras informações sobre a competência dos candidatos são utilizadas como base para a seleção daqueles que devem ingressar na universidade, além dos resultados das provas, interessa verificar o quanto estas últimas contribuem para aumentar a validade preditiva do conjunto de informações. No processo de seleção de candidatos aos "colleges" americanos, entre outras informações, são utilizadas as notas no curso secundário e o desempenho no *Scholastic Aptitude Test*; cabe perguntar qual é o incremento na predição das notas no "college", quando o teste de aptidão é empregado além das notas obtidas na escola secundária. O estudo de Fincher (1974), por exemplo, focaliza esta questão, analisando dados coletados durante treze anos com relação aos candidatos e ao respectivo comportamento na universidade. Outro tipo comum de estudo dá maior atenção à análise de regressão, estimando parâmetros de equações de regressão. Numa revisão da literatura de estudos de regressão que focalizam a validade de medidas utilizadas na seleção para o ingresso no "college", em diversas populações, Breland (1978) analisa os resultados de trinta e cinco estudos de regressão que utilizam diferentes preditores e diferentes critérios (ou comportamentos a serem previstos); entre os estudos brasileiros citados, o de Bessa e Mettel (1965) serve como exemplo

deste tipo investigação da validade preditiva de escores obtidos no processo de seleção para a universidade. Um outro exemplo é apresentado a seguir, utilizando escores obtidos em algumas provas da Fundação Cesgranrio e notas em matérias do primeiro ano de Engenharia da PUC-RJ.

Para este exemplo, tomaram-se escores (número de questões respondidas corretamente) obtidos nas provas de Português, Física, Matemática, Química e Biologia, pelos candidatos que se apresentaram ao vestibular da Fundação Cesgranrio em 1977 e ingressaram nos cursos de Engenharia da PUC-RJ. Foram também tomadas as notas dos alunos desse grupo que se matricularam efetivamente na PUC-RJ e que fizeram as provas de Física, Cálculo I e de Química no primeiro ano do curso<sup>3</sup>. Segundo informação do DAR, essas disciplinas são comuns a todos os cursos de Engenharia da PUC-RJ no primeiro ano. No total, inclui-se no trabalho o grupo de 256 alunos de Engenharia que apresentavam notas nas disciplinas consideradas. Adotou-se um modelo de análise de regressão múltipla multivariada, com três variáveis cujas médias interessava prever e com cinco variáveis preditoras — respectivamente, as notas de Física, Cálculo I e Química e as notas de Português, Física, Matemática, Química e Biologia do vestibular. Na especificação do modelo de regressão linear, as variáveis preditoras são aleatórias e os valores observados estão representados na matriz X; as variáveis cujas médias se deseja prever são também aleatórias e estão representadas na matriz Y. Assim, o modelo é representado pela equação

$$Y = XB + E, \text{ onde}$$

Y e X são matrizes que contêm, respectivamente, as notas nas provas do curso de Engenharia e as notas nas provas do vestibular; B é a matriz de coeficientes de regressão e E é a matriz de erros aleatórios, independentes, cuja expectância é igual a zero e cujas variâncias são constantes em relação a cada um dos valores de Y. A Tabela 2 apresenta os valores dos coeficientes de regressão padronizados, obtidos para o grupo de alunos estudado.

**TABELA 2**

**Coefficientes de Regressão Padronizados. Regressão das Notas de Física, de Cálculo I e de Química, sobre as Notas do Vestibular em Português, Física, Matemática, Química e Biologia (N = 256)**

Provas de Vestibular	Disciplinas do curso		
	Física	Cálculo I	Química
Português	.1342	.1068	.2184
Física	.1592	.0804	.0952
Matemática	.3265	.3539	.1233
Química	.1266	.0831	.2445
Biologia	.1526	.1051	.0856

Nota — Devido a erros detectados na perfuração de cartões, foi preciso processar os dados novamente. Por esta razão, os valores da Tabela 2 não coincidem com aqueles apresentados durante a palestra realizada na PUC-RJ.

(<sup>3</sup>) Cabe agradecer ao DAR, da PUC-RJ e, especialmente, à Sra. Maria Eugênia M. de Oliveira, que facilitaram o acesso às notas dos alunos. Para o processamento dos dados, o trabalho contou com o equipamento da RDC da PUC-RJ.

Tomando-se o grupo de alunos, considerado neste exemplo como a população de interesse para o pesquisador, os valores da Tabela 2 podem ser comparados entre si no contexto das variáveis introduzidas na equação de regressão, já que estão expressos em termos de escores padronizados. Assim, considerando-se o conjunto de variáveis preditivas introduzidas na equação, o aumento da nota média, na prova de Física do curso de Engenharia é maior, por unidade de aumento no escore de Matemática do teste do vestibular do que por unidade de aumento do escore das provas de Física, ou de Português, ou de Química, ou de Biologia. Também o aumento da nota média na prova de Cálculo I é maior, por unidade de aumento na prova de Matemática do vestibular do que por unidade de aumento no escore de qualquer das outras provas utilizadas na seleção. Já em relação à prova de Química do curso de Engenharia, a situação é diferente, sendo Português e Química, os melhores preditores; o aumento na nota média da prova de Química do curso de Engenharia é maior, por unidade de aumento do escore de Química e do escore de Português do vestibular do que por unidade de aumento do escore em Matemática, ou em Física ou em Biologia do mesmo vestibular.

Neste exemplo particular, verifica-se que o desempenho na prova de Matemática, utilizada no vestibular de 1977 da Fundação Cesgranrio, é a melhor preditor da nota média das provas de Física e da prova de Cálculo, no primeiro ano do curso de Engenharia da PUC-RJ; quanto à prova de Química desse mesmo ano, a nota média é melhor prevista pelos escores obtidos na prova de Química e na de Português do vestibular. Em relação a nossos vestibulares, um estudo deste tipo vale como verificação, a posteriori, da validade preditiva das provas utilizadas no vestibular, em relação a certas notas do curso tomadas como critério. A utilização dos resultados do estudo é bastante limitada, posto que as provas dos vestibulares nunca são repetidas, de modo que a equação de previsão não pode ser empregada para fazer-se uma previsão dos escores que um grupo de candidatos teria no curso de Engenharia da PUC-RJ, com base nos escores obtidos nas provas do vestibular.

Por várias razões, os estudos de validade preditiva dos escores obtidos em vestibulares brasileiros devem ser encarados com bastante cautela, apesar de serem grandemente necessários para um melhor conhecimento tanto dos testes empregados na seleção como das provas de rendimento nas universidades. Em geral esses estudos só podem ser realizados com base em escores de grupos já selecionados pelo vestibular — não há possibilidade, na prática, de serem admitidos, a título experimental, candidatos que não são classificados no vestibular. De outro lado, dificilmente as distribuições de notas obtidas nas provas dos cursos universitários são tais que se justifique o pressuposto de uma distribuição multivariada normal. Assim sendo, nem sempre as correlações obtidas podem ser justamente interpretadas. Também, nem sempre são justificáveis os pressupostos que devem ser introduzidos para interpretação de testes de significância, quando se selecionam amostras de candidatos que prestam exames vestibulares. Do outro lado, as conclusões sobre a contribuição do desempenho em cada prova utilizada na seleção dos candidatos devem limitar-se ao contexto das provas em que a análise se baseia; ou seja, a interpretação dos coeficientes de regressão limita-se ao efeito de cada variável no conjunto das variáveis introduzidas na equação de regressão. Se o conjunto de provas empregadas na previsão dos resultados dos alunos nos cursos universitários se altera, os coeficientes de regressão devem ser obtidos e interpretados dentro do novo contexto de variáveis preditivas.

Do exemplo e das considerações sobre a interpretação dos resultados apresentados na Tabela 2, conclui-se que o estudo da validade preditiva, nas condições em que se realizam os vestibulares brasileiros, tem sua utilidade prática limitada pelo fato das provas não poderem ser repetidas em várias ocasiões. Ao contrário do que acontece nos casos em que testes padronizados são empregados na seleção de candidatos ao ingresso nos cursos universitários, nos vestibulares brasileiros as equações de regressão não podem ser utilizadas na previsão dos resultados dos alunos que se candidatam em ocasiões diferentes.

## O APERFEIÇOAMENTO DO PROCESSO DE SELEÇÃO

A única maneira de encontrar soluções para os problemas da seleção para o ingresso nas universidades brasileiras é produzir o máximo de pesquisas empíricas e de avaliações, observando os requisitos técnicos para que as conclusões tenham a validade. Pretender resolver esses problemas sem base em estudos empíricos, é façanha que não se justifica atualmente.

Note-se que uma das dificuldades para a realização de pesquisas metodológicas, na área educacional, no Brasil, é a extrema escassez de pessoal especializado. Poucos são os profissionais formados no exterior e raros são os casos em que essa formação está sendo realizada no Brasil. De qualquer modo, os trabalhos começam a surgir agora, embora atingindo um público limitado pela falta de pessoal especializado.

A pesquisa educacional requer tempo, além de recursos financeiros — tempo para planejamento e implementação, tempo para divulgação, tempo para multiplicação de réplicas da pesquisa, tempo para assimilação dos resultados por aqueles que deles necessitam para aplicações práticas. De outro lado, aqueles que sentem os problemas, na prática, e experimentem a necessidade de introduzir alterações no sistema educacional reclamam soluções com urgência. Não é fácil para a comunidade de educadores aprender a esperar por soluções fundamentadas nos resultados da pesquisa educacional e a apreciar devidamente a resposta dos pesquisadores, quando estes simplesmente dizem que não conhecem suficientemente a questão para oferecer resposta definida. Comentando a história da pesquisa educacional nos Estados Unidos, Schutz (1979) opina que em relação a alguns setores da educação americana, os pesquisadores já podem ter respostas que serão de algum auxílio na solução de problemas práticos, mas em outros setores a resposta dos pesquisadores educacionais continua sendo “não sabemos o bastante”.

Terá a pesquisa educacional, no Brasil, respostas para os problemas colocados pelos educadores e pelo público em geral, em relação ao processo de seleção para o ingresso nas universidades? Os problemas exigem solução; se os pesquisadores são cautelosos nas respostas, não há como adiar as pesquisas que poderão fundamentar o desenvolvimento de processos de seleção em bases objetivas. As alterações que se introduzem em um sistema educacional têm um custo, que precisa ser medido e precisa ter avaliados os benefícios correspondentes. Essa é uma das razões para serem precedidas de pesquisas, que façam prever os resultados que deverão produzir. As modificações no processo de seleção para as universidades constituem apenas um caso particular. Não se justifica, pois, que as alterações nos exames vestibulares continuem sendo feitas com base na retórica ou na experiência de cada um de nós simplesmente.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANASTASI, Anne. *Psychological testing* (2nd ed.). New York: Macmillan, 1961.
- BARROSO, Carmen L. de M. Validade de conteúdo e preditiva das provas. *Ciência e Cultura*, 1970, 22(3), 60-67.
- BARROSO, Carmen L. de M. Estudos de predição do comportamento acadêmico II. *Cadernos de Pesquisa*, 1973, 5, 55-76.
- BARROSO, C.L. de M., RIBEIRO Netto, A. COELHO, Maria Helena M.C. Estudos de predição do comportamento acadêmico I. *Cadernos de Pesquisa*, 1973, 5, 37-54.
- BEATON, A.E., HILTON, T.L., SCHRADER, W.B. *Changes in the verbal abilities of high school seniors, college entrants, and SAT candidates between 1960 and 1972*. Princeton: Educational Testing Service, 1977.
- BESSA, N.M., METTEL, T.L. Validade de três testes do DAT (Forma B). *Arquivos Brasileiros de Psicotécnica*, 1965, 17(3), 5-15.
- BESSA, N.M. Exames vestibulares e problemas de padronização de testes. *Ciência e Cultura*, 1970, 22(3), 54-59.

- BESSA, N.M. *Testes de desenvolvimento educacional-Manual*. Rio de Janeiro: Fundação Getúlio Vargas, 1971.
- BESSA, N.M. O poder preditivo dos testes de aptidão profissional. *Orientação – DOEP*, 1973, 11, 43-57.
- BESSA, N.M. *University entrance examinations in Brazil: Measurement procedures*. São Paulo: Fundação Carlos Chagas, 1976.
- BLOOM, B.S., ENGLEHART, M.D., FURST, E.J., HILL, W.H., KRATHWOHL, D.R. *Taxonomy of educational objectives. Handbook I; Cognitive domain*. New York: Longmans, Green, 1956.
- BRAY, D.W., MOSES, J.L. Personnel selection. *Annual Review of Psychology*, 1972, 23, 545-576.
- BRELAND, Hunter M. *Population validity and college entrance measures*. Princeton: ETS/CEEB Research and Development Report – (RDR 78-79, 2), 1978.
- BUCHWEITZ, Bernardo. Testes de múltipla escolha e de resposta livre em Física Geral. *CADERNOS de Pesquisa*, 1976, 16, 3-6.
- BUROS, O.K. Fifty years in testing: Some reminiscences, criticisms, and suggestions. *Educational Researches*, 1977, 6(7), 9-10.
- BURT, Cyril. *Mental and scholastic tests* (4th ed.). Londres: Staple Press, 1962.
- CHEDIAK, A.J., BESSA, N.M., BECHARA, E.C. et alii. *Parecer sobre a inclusão da prova discursiva em concursos vestibulares*. Rio de Janeiro: Fundação Cesgranrio, 1975, Mimeografado.
- COULSON, John E. National evaluation of the Emergency School Act (ESAA): A review of methodological issues. *Journal of Educational Statistics*, 1978, 3(1), 1-60.
- CRONBACH, L.J., WARRINGTON, W.G. Efficiency of multiple-choice test as a function of spread of item difficulties. In W.A. MEHRENS, R.L. EBEL (Eds.), *Principles of educational and psychological measurement*. Chicago: Rand McNally, 1967, 404-425.
- DAVIS, Frederick B. (Chair). *Standards for educational and psychological tests*. Washington: American Psychological Association, 1974.
- DUBOIS, Phillip H. A test dominated society: China, 1115 B.C. – 1905 A.D. In A. ANASTASI (Ed.), *Testing problems in perspective*. Washington: American Council on Education, 1966, 29-36.
- DUNCAN, O.D., FEATHERMAN, D.L., DUNCAN Beverly. *Socioeconomic background and achievement*. New York: Seminar Press, 1972.
- EDUCATIONAL TESTING SERVICE. *Testing programs. Special Services. Instructional activities*. Princeton: ETS, 1968.
- FERNANDES, Lúcia M. Características técnicas dos instrumentos de medida. *Ciência e Cultura*, 1970, 22(3), 50-59.
- FESHBACH, S. The illusory control group in field research: Comments on the ESAA evaluation. *Journal of Educational Statistics*, 1978, 3(1), 87-93.
- FINCHER, Cameron. Is the SAT worth its salt? An evaluation of the Scholastic Aptitude Test in the university system of Georgia over a thirteen-year period. *Review of Educational Research*, 1974, 44(3), 293-305.
- GUILFORD, J.P. *Psychometric methods* (2nd. ed.). New York: McGraw-Hill, 1954.
- GULLIKSEN, H. *Theory of mental tests*. New York: John Wiley, 1950.
- JAMESON, S.C., OTTOBRE, F.M. United States. In F.M. Ottobre, *Admission to higher education in sixteen countries and some international developments*. Princeton: International Association for Educational Assessment, 1976.
- KENNEDY, Mary M. Findings from the Follow Through Planned Variation Study. *Educational Researcher*, 1978, 7(6), 3-11.
- LORD, F.M., NOVICK, M.R. *Statistical theory of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.
- Ministério da Educação e Cultura, Departamento de Assuntos Universitários. *Pesquisa sobre demanda no ensino superior*. Brasília: Ministério da Educação e Cultura, 1974.

- MONTEIRO, Kilda N. Estudo com o D.A.T. (Forma A). *Arquivos Brasileiros de Psicotécnica*, 1964, 16(4), 47-54.
- MYERS, A.E., MCCONVILLE, C.B., COFFMAN, V.E. Simplex structure in the grading of essay tests. *Educational and Psychological Measurement*, 1966, 26, 41-54.
- OTTOBRE, F.M. (Ed). *Admission to higher education in sixteen countries and some international developments*. Princeton: International Association for Educational Assessment, 1976.
- PETTIGREW, T.F. Competing evaluation models: The ESAA evaluation. *Journal of Educational Statistics*, 1978, 3(1), 99-106.
- ROGOSA, D. Politics, process, and pyramids. *Journal of Educational Statistics*. 1978, 3(1), 79-86.
- SANCHEZ, A.M., CARONE, F.B., VIANNA, H.M., LAURITO, I.B.G., MORAES, C.D. *Validade de conteúdo de uma prova de comunicação e expressão; análise de alguns problemas*, 1979, Mimeografado.
- SCHUTZ, R.E. Where we've been, where we are, and where we're going in Educational R&D. *Educational Researcher*, 1979, 8(8), 6-8.
- SCRIVEN, M. The methodology of evaluation. *AERA Monograph Series on Curriculum Evaluation*, 1. Chicago: Rand McNally, 1967, 39-83.
- SEDDON, G.M. The properties of Bloom's Taxonomy of Educational Objectives for the Cognitive Domain. *Review of Educational Research*, 1978, 48(2), 303-323.
- SOARES, Magda B., A redação no vestibular. *Cadernos de Pesquisa*. 1978, 24, 53-56.
- STUFFLEBEAM, D.L., FOLEY, W.J., GEPHART, W.J., GUBA, E.G., HAMMOND, R.L., MERRIMAN, H.O., PROVUS, M.M. *Educational evaluation and decision making*. Itasca, Illinois: F.E. Peacock, 1971.
- The Evaluation Center. Standards for evaluation of educational programs, projects, and materials. *Educational Researcher*, 1979, 8(1), 26.
- VIANNA, Heraldo M. Emprego e características de provas objetivas. *Ciência e Cultura*, 1970, 22(3), 68-74.
- VIANNA, Heraldo M. *Avaliação das provas do concurso vestibular da área biomédica*. São Paulo: Fundação Carlos Chagas, Mimeografado. 1974.
- VIANNA, Heraldo M. Redação e medida da expressão escrita: algumas contribuições da pesquisa educacional. *Cadernos de Pesquisa*, 1976, 16, 41-47.
- VIANNA, Heraldo M. Flutuações de julgamento em provas de redação. *Cadernos de Pesquisa*, 1976, 19, 5-9.
- VIANNA, Heraldo M. Aplicação de critérios de correção em provas de redação. *Cadernos de Pesquisa*, 1978, 26, 29-34.
- WEISBERG, Herbert I. How much does ESAA really accelerate academic growth? *Journal of Educational Statistics*, 1978, 3(1), 69-78.