



**Centre pour le développement de tests et le diagnostic  
à l'Institut de psychologie de l'Université de Fribourg**

---

# **L'examen d'aptitude aux études de médecine en Suisse**

**Critères et procédés d'application du test**

Rapport sur le Symposium international à Berne  
Editeurs: K.-D. Hänsgen; N. Ischi

---

**Rapport 3 (1997)**

---

## Table des matières

|   |     |
|---|-----|
| Table des matières .....  | 2   |
| Avant-propos des éditeurs .....   | 3   |
| Critères d'aptitudes pour l'admission aux études de médecine en Europe:<br>résultats d'une enquête menée dans tous les pays d'Europe<br><i>Günter Trost</i> .....                       | 6   |
| The Swedish Scholastic Aptitude. Test Research and main findings<br><i>Ingemar Wedman &amp; Widar Henriksson</i> .....  | 21  |
| The Medical College Admission Test (MCAT) - its use in U.S.<br>and Canada and some results of validation studies<br><i>John L. Hackett</i> .....  | 42  |
| Admission to the study in medicine in Belgium: two 'different' solutions<br>to the 'same' problem; reflections of a Flemish school psychologist<br><i>Piet J. Janssen</i> .....         | 57  |
| Le test d'aptitude aux études de médecine en Suisse en tant qu'instrument<br>de limitation de l'accès aux études<br><i>Klaus-Dieter Hänsgen</i> .....                                   | 70  |
| Test d'aptitude aux études de médecine en Suisse: dans quelle mesure<br>peut-on s'y entraîner?<br><i>Rainer Hofer &amp; Klaus-Dieter Hänsgen</i> .....                                  | 83  |
| Qualités requises d'une procédure d'admission aux études de médecine en Suisse:<br>lignes directrices pour la mise au point d'un diagnostic d'aptitude<br><i>Urs Schallberger</i> ..... | 91  |
| Adresses.....   | 100 |

---

Rédaction: Sébastien Simonet

© Centre pour développement de tests et le diagnostic à l'Institut de Psychologie de  
l'Université de Fribourg, Route d'Englisberg 9, CH-1763 Granges-Paccot  
Tél. 026/300 79 86 - Fax 026/300 97 63 - e-mail: ZTD@UNIFR.CH  
Internet: <http://www.unifr.ch/Main/ztd.html>

## Avant-propos des éditeurs

Dans tous les pays industrialisés et hautement développés, la demande de formation universitaire ne cesse de s'accroître; les études de médecine, notamment, sont très prisées. Il est de pratique courante dans nombre de pays de limiter l'admission aux études de médecine par un *numerus clausus* ajusté aux capacités d'accueil, visant à maintenir le bon fonctionnement des universités et la qualité de la formation. Mais toute restriction nécessite des bases légales suffisantes ainsi que des critères d'admission scientifiquement éprouvés, équitables et reconnus.

On sait qu'en Suisse, depuis plusieurs années, avec des conditions d'accès à l'université jusqu'ici illimitées, les universités ont tendance à répondre au grand intérêt témoigné aux études de médecine par un dépassement des capacités de formation. Or, un tiers environ des candidats admis, abandonnent au cours des deux premières années - en partie en raison d'un renforcement de la sélection intra-universitaire en tant que "mesure d'urgence" prise pour adapter le nombre des étudiants aux capacités disponibles en formation clinique. Cette capacité est limitée sans appel, notamment par le nombre de patients et de lits dans les hôpitaux.

La Suisse, c'est connu, figure parmi les pays ayant la plus haute densité de places d'études en médecine et, en conséquence, la plus forte densité médicale. Etant donné que la part extrêmement élevée des ressources financières attribuée aux facultés de médecine ne peut plus être augmentée, il est impératif d'endiguer à temps une explosion des coûts de la santé et d'utiliser de manière rationnelle les moyens des universités; les milieux politiques et universitaires évoquent périodiquement l'opportunité d'un *numerus clausus* en tant que mesure d'urgence. Enfin - et ce n'est pas le moindre des arguments - les universités et les étudiants eux-mêmes déplorent la baisse de qualité de la formation et reconnaissent la nécessité et l'urgence d'une réforme des études de médecine. Toutefois, les réformes déjà introduites au sein des universités de Genève, Lausanne et Berne exigent une réduction supplémentaire des capacités d'accueil.

Etant donné que les cantons et la Confédération dépensent en moyenne quelques 60'000 francs par an pour la formation d'un étudiant en médecine pendant les deux premières années de ses études, on ne peut que juger inacceptable, pour des raisons économiques également, le nombre élevé d'interruption des études. A condition que les universités appliquent le **critère de l'aptitude aux études** et règlent le nombre d'admissions d'après la capacité d'accueil, la probabilité d'une interruption des études sera beaucoup plus faible et les investissements atteindront leur but.

En 1995 pour la première fois, des préparatifs ont été mis en route à l'échelle nationale en vue d'introduire un test d'aptitude en tant que critère d'admission. Pour diverses raisons - entre autres l'absence de bases légales et des changements de calendrier dans l'année scolaire de certains gymnases - il a fallu différer une fois de plus la mise sur pied du test. Cependant, le problème de surcharge demeure et rien n'indique qu'il se résoudra spontanément. De plus, une formule de rechange au test d'aptitude a fait son apparition, qui aurait la préférence de certains cantons, à savoir une sélection renforcée au cours de la première année d'études. Mais des solutions judicieuses et durables au regard de la politique universitaire ne peuvent être menées à bien que par un travail sur le plan national. L'année passée déjà, mais cette année plus sûrement encore, il ne sera pas simple pour les candidats, dans le choix de l'université où ils voudraient étudier la médecine, de peser les avantages et les inconvénients des conditions d'études.

Le problème ne pouvant se résoudre qu'à l'échelle nationale, il est important de connaître les moyens qui, à l'étranger, se sont révélés efficaces dans la régulation du nombre d'admissions. De nombreux pays ont accumulé dans ce domaine de nombreuses expériences dont nous devons tirer profit. C'est pourquoi un symposium international a eu lieu le 8 novembre 1996 à Berne sur le thème "Examen d'aptitude aux études de médecine - critères et procédés d'application de tests".

Les exposés présentés par huit invités ont suscité une discussion nourrie parmi les représentants des lycées, des universités, des cliniques universitaires, des conseillers aux études universitaires, des associations professionnelles de la médecine ainsi que parmi les représentants des organes universitaires et de la politique en matière de santé.

G. Trost apporta les **résultats d'une enquête** qui devait établir quels **pays d'Europe** appliqueraient des mesures, et lesquelles, pour régler l'admission aux études de médecine. Il n'y a plus guère de pays qui puissent se payer le luxe de ne rien faire. On recourt très souvent à des tests standardisés fondés sur des critères qui doivent poser de hautes exigences.

Le présent rapport est consacré à deux concepts de tests pratiqués: J. Hackett présente le **MCAT (Medical College Admission Test)** qui a cours aux Etats-Unis et au Canada, et I. Wedmann le test suédois SweSAT. Les deux tests, tout comme le test allemand pour les études de médecine (TMS), peuvent être considérés comme des instruments scientifiquement éprouvés qui ont aussi trouvé l'adhésion nécessaire dans les pays où ils sont appliqués.

P.J. Janssen rapporte sur **l'introduction d'un test en Belgique**, pays qui par son plurilinguisme présente des conditions générales comparables à la Suisse.

Les rapports de U. Schallberger et K.-D. Hänsgen abordent la situation en Suisse: rappel des démarches déjà entreprises et définition une fois de plus des exigences auxquelles doit satisfaire une procédure d'admission.

En marge du programme du symposium, ce rapport contient une contribution de R. Hofer et K.-D. Hänsgen sur la question de savoir s'il est possible de **s'entraîner au test**. Si le test est introduit en Suisse, il faudra probablement se demander si les offres d'entraînement de diverses entreprises privées pourraient entraver l'égalité des chances. Il apparaît dès lors urgent de se servir des résultats de la présente étude pour opposer un point de vue critique et très dissuasif aux promesses de telles offres.

De l'avis des organisateurs et des auteurs de ce rapport, un test d'aptitude tenant compte des conditions propres à la Suisse représente en définitive la méthode de choix pour l'admission aux études de médecine. Le test à utiliser est scientifiquement éprouvé et garanti; il existe sous trois formes linguistiques équivalentes. L'égalité des chances est totalement assurée notamment par une exécution et une évaluation standardisées. Puissent les arguments et faits ici présentés faire avancer la discussion en Suisse!

K.-D. Hänsgen

N. Ischi

Fribourg/Berne, en décembre 1996

# **Critères d'aptitudes pour l'admission aux études de médecine en Europe: résultats d'une enquête menée dans tous les pays d'Europe**

**G. Trost**

## **1. Buts et historique de l'enquête**

En 1995, Judith Ebach et moi-même avons mené dans tous les pays d'Europe une enquête sur l'admission aux études de médecine à l'instigation de l'Association of Medical Schools of Europe (AMSE). Un groupe de travail de cette association, présidé par le professeur Sergio Curtoni, Turin, nous a secondé dans notre recherche d'interlocuteurs compétents (en règle générale des doyens des facultés de médecine ou des représentants d'associations nationales de professeurs de médecine des différents pays). L'enquête a été encouragée par la Commission européenne dans le cadre du programme ERASMUS.

L'enquête visait deux objectifs:

- a) La tâche consistait à établir et à exposer avec clarté, pour chaque pays, les modes de procédure, les critères de sélection et les instruments de diagnostic utilisés lors de l'admission aux établissements universitaires de formation médicale.
- b) Il s'agissait de recueillir et de mettre en valeur dans une analyse supra-nationale les expériences faites en matière de critères d'admission et d'instruments de sélection dans les différents pays et les résultats d'études empiriques d'évaluation.

Grâce à une série de campagnes de rappel, nous avons finalement été en possession d'informations de 35 pays d'Europe dotés d'établissements universitaires de formation médicale.

Dans la présente contribution, nous donnerons un aperçu des critères appliqués dans les pays européens pour décider de l'admission aux études de médecine. En outre, les résultats d'études empiriques d'évaluation des instruments de sélection les plus importants seront exposés en version abrégée. La version complète de l'étude paraîtra ces jours sous forme de livre (Ebach & Trost, 1996).

## **2. Trois modèles de base de l'admission aux études universitaires**

Tout d'abord nous exposerons trois modèles de base, constituant trois catégories de réglementations nationales d'admission.

Selon le modèle de base I, chaque candidat en possession de la pièce justificative formelle d'admission aux études universitaires (c'est en général le certificat de fin d'études du degré secondaire) peut accéder immédiatement aux études universitaires. Si le nombre de candidats dépasse le nombre de places disponibles dans les curricula concernés pour un temps assez long, la sélection s'opère durant les études, la plupart du temps après la première année d'études, plus rarement après la deuxième année d'études, sur la base d'examens universitaires. La sélection peut être à un palier ou à plusieurs paliers.

Font partie des représentants de ce modèle de base la France, l'Autriche et, du moins cette année encore, la Suisse. La Belgique faisait elle jusqu'ici aussi partie de cette catégorie. Mais à partir de l'année prochaine, la partie flamande du pays introduira une procédure de test avant le début des études pour sélectionner les candidats aux études de médecine (cf. exposé de Piet Janssen dans le présent rapport).

Selon les modèles de base II et III, la sélection se fait avant l'admission.

Le modèle de base II repose sur l'hypothèse que les candidats ayant réussi les études scolaires du degré secondaire sont par principe aptes à faire des études universitaires. En cas d'excédent de candidats, il est procédé à un tirage au sort, c'est-à-dire que la sélection est fortuite. En réalité, ce principe n'est toutefois appliqué nulle part de manière tout à fait stricte. En République fédérale d'Allemagne, une "procédure de tirage au sort affinée d'un facteur de performances", fondée sur le principe que les chances offertes par le tirage au sort étaient codéterminées par la hauteur de la note moyenne obtenue à la maturité, faisait partie intégrante de la procédure de passage pour l'admission aux curricula médicaux durant la période de 1980 à 1985. Aux Pays-Bas on a recours, dans les disciplines dites de *numerus fixus* à une procédure échelonnée selon laquelle la sélection se fait aussi bien d'après la performance que d'après le hasard: le premier échelon est la sélection selon la note moyenne obtenue à l'école; le deuxième échelon est soumis à une procédure de tirage au sort.

La grande majorité des pays, à savoir 31 des 35 pays européens, de même que certains pays non européens, appliquent, pour l'admission aux études de médecine, le modèle de base III: dans ces pays, l'admission se fait selon les critères des aptitudes et des performances; ces dernières peuvent être établies de diverses manières avant le début des études. A l'intérieur de ce modèle, on distingue des modes centralisés et des modes décentralisés de procédure. Dans le premier cas, les critères d'admission sont fixés selon des normes uniformes; ils sont dès lors contraignants pour toutes les universités, comme c'est le cas pour les curricula à *numerus clausus* en République fédérale d'Allemagne. Dans le deuxième cas, les universités fixent chacune les critères d'admission et

statuent de manière autonome; à titre d'exemple de ce mode de sélection, nous citerons la pratique en usage en Grande-Bretagne.

Dans un premier résumé, on peut conclure ce qui suit: pratiquement partout en Europe une sélection est opérée entre les candidats aux études de médecine. Dans la plupart des pays, la décision de sélection est prise avant l'admission aux études.

Dans le chapitre suivant, on considérera les pays d'Europe dans lesquels une sélection a lieu avant le début des études. La question qui retiendra notre attention dans ce contexte est de savoir sur la base de quels critères la sélection est opérée.

### **3. Critères de sélection appliqués pour l'admission aux établissements universitaires de formation médicale en Europe**

Si l'on examine pour commencer le *degré d'uniformité* des procédures d'admission dans les différents pays d'Europe, on constate que dans douze pays les mêmes critères sont valables dans tous les établissements universitaires de formation médicale; dans neuf pays, les critères sont similaires dans la plupart des établissements universitaires de formation médicale, dans trois pays (la Bulgarie, la Suède et la Tchéquie) ils diffèrent d'une université à l'autre. Six autres pays n'ont qu'un établissement universitaire de formation médicale. (Pour la Roumanie, nous n'avons pas d'indication à ce sujet.)

Le critère le plus fréquent de sélection est la *performance qui figure dans le certificat de fin de scolarité secondaire*. Dans 23 pays d'Europe, cet indicateur des aptitudes aux études est pris en considération pour la décision d'admission. Dans douze pays, certaines disciplines sont exigées au degré secondaire, généralement les sciences naturelles (tableau 1).

Le critère qui vient en deuxième ordre de fréquence est le résultat obtenu lors de *tests de connaissances* portant sur des matières scolaires.

Dans 22 pays d'Europe, les résultats de ce genre de tests jouent un rôle pour l'admission aux études de médecine. Généralement les contenus des branches scolaires biologie, chimie et physique font l'objet de l'examen par les tests; dans chaque quatrième de ces pays environ il est tenu compte, en outre, des résultats obtenus en mathématiques ou en statistique, de la manière dont les candidats s'expriment dans leur langue maternelle ou des résultats obtenus pour un test en langues étrangères. Les examens portent non seulement sur des connaissances pures, mais aussi en partie sur une compréhension générale des domaines thématiques.



Dans 18 pays, ces tests de connaissances se présentent, du moins en partie, sous forme de devoirs revêtant la forme de questions à choix multiple (Multiple-Choice-Items). La durée des tests varie considérablement; elle va d'une heure (dans certaines universités italiennes) à 19 heures (en Espagne).

A la différence des tests de connaissances portant sur des matières scolaires, *les tests d'aptitudes aux études* n'ont pas pour but de mesurer les connaissances acquises à l'école, mais d'appréhender des aptitudes cognitives particulièrement importantes pour la maîtrise des exigences posées par les curricula en question. De telles procédures de test sont utilisées dans cinq pays européens. En Allemagne, tous les candidats doivent se soumettre, à l'heure actuelle, à un test spécifique d'aptitudes aux études, appelé Test pour les études de médecine, pour l'obtention d'une place d'études de médecine. En Turquie, un test général d'aptitudes aux études représente le premier échelon d'un programme de test à deux paliers; le deuxième palier consiste en tests de connaissances fondés sur des domaines spécifiques. En Suède, les candidats peuvent dire s'ils préfèrent se soumettre à un test d'aptitudes aux études ou s'ils veulent être classés selon leur certificat de fin d'études secondaires pour être admis aux études de médecine (cf. exposé d'Ingemar Wedman dans le présent rapport). En Finlande et en Tchéquie, certaines universités utilisent les résultats de tests d'aptitudes aux études comme critères d'admission.

A l'exception du test finlandais, les tests d'aptitudes aux études utilisés en Europe comportent uniquement des devoirs avec des questions à réponses multiples. La durée est variable. Le test suédois par exemple nécessite un peu plus de quatre heures, le test allemand cinq heures.

Les résultats *d'interviews* sont pris en considération dans sept pays d'Europe, pour la décision de sélection, mais pas dans tous les établissements universitaires de formation médicale (Danemark, Allemagne, Grande-Bretagne, Suède, Tchéquie, Ukraine, Biélorussie). Le nombre de personnes participant à une interview varie de pays à pays et en partie d'université à université. En Grande-Bretagne, le candidat a en face de lui parfois cinq interviewers; en Allemagne un candidat doit toujours faire face à deux interlocuteurs, à l'Institut Karolinska à Stockholm deux entretiens ont lieu avec, à chaque fois, un interviewer.

Aucun pays d'Europe n'exige de manière contraignante un *stage hospitalier* comme critère d'admission. Au Danemark et en Biélorussie, les candidats peuvent cependant améliorer leurs chances s'ils justifient d'une telle expérience pratique. En Norvège, cela vaut pour toute forme d'expérience professionnelle. En Allemagne, les candidats peuvent améliorer leur position sur la liste d'attente en acquérant une formation professionnelle et en effectuant un service social d'au moins une année durant le "temps d'attente".

Les critères suivants ne sont pris en considération que dans un pays européen: le *temps d'attente* (Allemagne), une *lettre de recommandation* émanant de la direction de l'école ou "*Statement of Interest*", que se délivrent les candidats eux-mêmes (Grande-Bretagne), un prix reçu lors d'un *concours de sciences naturelles* (Croatie), le résultat d'un *tirage au sort* (Pays-Bas), un certificat de "*capacité à la communication interpersonnelle*" (Portugal).

De tous les pays d'Europe dans lesquels ont été décrétées des limitations d'admission, il n'y en a qu'un (la Slovénie), dans lequel la sélection se fait sur la base d'un seul critère: les notes du degré secondaire. Cette réglementation est valable depuis 1995; auparavant, on tenait compte aussi des résultats obtenus lors de tests de connaissances. Dans tous les autres pays, les étudiants sont sélectionnés sur la base d'une *combinaison de plusieurs critères*. A titre d'exemple d'un système d'admission particulièrement différencié, où dans un modèle de quotas quatre critères différents sont pris en compte, nous brosserons un bref tableau de la procédure de sélection appliquée en Allemagne pour les études de médecine:<sup>1</sup>

- 45 pour cent des places sont attribués à des candidats qui présentent le nombre de points le plus élevé, calculés en combinant la note moyenne de maturité et le résultat du Test pour les études de médecine selon un rapport de 55 à 45;
- 10 pour cent des places sont attribués à des candidats qui ont obtenu les meilleurs résultats de test sans prise en considération de leur note moyenne de maturité.
- 20 pour cent des places sont attribués à des candidats qui présentent le plus long temps d'attente;
- 15 pour cent des places sont attribués sur la base des résultats d'entretiens de sélection menés par les universités;
- 10 pour cent des places sont prévus pour des groupes déterminés de candidats tels que les candidats à une deuxième filière d'études et pour des cas de rigueur (Trost, 1989).

La *combinaison de loin la plus fréquente* de critères d'admission est celle des notes scolaires et des *tests de connaissances fondés sur des matières scolaires*.

---

<sup>1</sup>En République fédérale d'Allemagne la procédure de sélection pour les études de médecine ici évoquée sera remplacée au semestre d'été 1998 par une procédure simplifiée. Les Etats fédérés ont certes décidé de s'en tenir aux restrictions d'admission - face à des mesures d'économie draconiennes, une majorité de „Länder“ s'est toutefois prononcée pour une procédure de sélection moins coûteuse, fondée sur les deux critères: note moyenne de maturité et temps d'attente.

On trouve cette combinaison dans 15 pays européens; dans six de ces pays, la sélection se fait encore selon d'autres critères.

#### 4. Evaluation des instruments de sélection

Une analyse des procédures appliquées pour l'admission aux études universitaires dans les pays d'Europe serait insatisfaisante si l'on se bornait à décrire les pratiques actuelles. Hormis les réponses à la question: "Que se passe-t-il dans les différents pays?" les réponses à une deuxième question sont intéressantes. "Quelle est la qualité de fonctionnement de ce qui se fait dans les divers pays?" C'est pourquoi lors de notre enquête, nous avons prié les informants dans les divers pays de donner une appréciation personnelle de la procédure d'admission en vigueur à cette époque et de fournir tous les types d'études pour l'évaluation des différentes procédures. En outre, la littérature spécialisée internationale montre des résultats probants quant au bien fondé des contrôles probatoires précédant l'admission aux études universitaires. Les résultats seront brièvement exposés dans ce chapitre.

Comme aide à l'appréciation des différents instruments de sélection, on peut se servir des critères de qualité classiques utilisés en diagnostic pédagogique:

L'*objectivité* d'un instrument de sélection désigne le degré d'égalité de traitement qui est celui des candidats en relation avec l'exécution et l'évaluation d'une procédure ainsi que l'interprétation de leurs réactions dans cette procédure de sélection.

La *fiabilité* d'une procédure de sélection dépend de sa capacité à saisir avec exactitude les aptitudes à vérifier. Une procédure fiable donne par exemple des résultats identiques lorsque les tests sont répétés.

La *validité*: un instrument de sélection est déclaré *valable* s'il permet réellement de mesurer ce qu'il doit mesurer. Dans ce contexte, la valeur prédictive du procédé est spécialement importante: les résultats de la procédure de sélection devraient permettre de prédire le succès dans les études concernées. La qualité des critères disponibles de réussite (ou non) - ici les indications de réussite des études - doit également être considérée lors de l'appréciation de la valeur prédictive d'un instrument de sélection. Un pronostic satisfaisant ne sera obtenu, quelle que soit la procédure de sélection, que si les critères de réussite sont choisis avec suffisamment de pertinence, c'est à dire s'ils sont valables et fiables.

L'utilité d'une procédure de sélection ne se mesure pas uniquement sur la base de critères scientifiques, mais également sur d'autres critères dont:

- *l'aspect économique* qui comprend l'économie réalisée grâce au test d'aptitude en évitant des abandons en cours d'étude ou des échecs dus à une mauvaise orientation. Ces économies doivent être comparées aux coûts de mise en place d'une procédure de sélection comprenant le développement du test, sa mise en œuvre et son exploitation. Les coûts d'une telle procédure sont eux plus faciles à chiffrer.
- *l'équité* de la procédure de sélection. Une procédure de sélection est équitable par rapport à des groupes déterminés de candidats si l'on donne aux groupes partiels, qui ont les mêmes perspectives de succès, les mêmes chances d'admission.
- *l'acceptabilité* de la procédure de sélection, c'est-à-dire dans quelle mesure cette procédure est acceptée par les candidats aux études, les universités, les instances politiques et administratives responsables de l'enseignement ainsi que par l'opinion publique.
- Enfin, il faut toujours vérifier les *effets* indésirables qu'une procédure de sélection peut avoir sur le *comportement des candidats* ou sur l'enseignement du niveau secondaire.

Jusqu'à présent peu d'enquêtes empiriques ont étudié l'impact de ces différents critères sur les diverses formes de procédures de sélection. Le jugement des procédures de sélection s'appuiera donc pour la suite de l'exposé uniquement sur les critères d'objectivité et de fiabilité, de validité prédictive, d'économie et d'acceptabilité. Pour plus de clarté, nous allons nous référer aux éléments de sélection les plus souvent utilisés: les notes scolaires, les tests de connaissances et de capacité et les interviews.

#### **4.1. Les notes scolaires**

Les notes sont en principe considérées comme valables pour mesurer un degré d'acquisitions scolaires atteint. Indirectement, elles mesurent aussi les capacités intellectuelles. Si les notes scolaires sont utilisées comme instrument de sélection pour l'entrée aux universités, nous pouvons penser qu'en plus de leur fonction diagnostique, elles auront d'autres fonctions, en particulier une fonction pédagogique.

Les notes scolaires font preuve uniquement d'une objectivité moyenne dans le sens d'une conformité de jugement et d'une faible fiabilité en cas de répétition des mêmes travaux (Baron-Bolt, 1989; Ingekamp, 1989). Suffisamment d'études démontrent les différences d'évaluation des performances scolaires suivant les enseignants, les écoles ou les pays.

Malgré toutes les faiblesses de ce système de notation, c'est le meilleur moyen de comparaison révélateur des capacités générales pendant les dernières années scolaires, permettant de prédire le succès ou non des études (Baron-Bolt, 1989; Trost, Klieme & Nauels, 1996).

La capacité de prévoir le succès dans les études au moyen des notes par matière peut certes varier sensiblement selon la branche, mais même les indices les plus élevés de validité pronostique obtenus sur la base des notes par matière restent nettement inférieurs à ceux obtenus sur la base des moyennes des notes (Baron-Boldt, 1989, Weingardt, 1989).

Du fait qu'elles sont facilement accessibles, les notes scolaires représentent un critère de sélection favorable sous l'aspect du coût.

En tant que critère de sélection pour l'admission à l'université, elles recueillent un degré d'acceptabilité moyen. En République fédérale d'Allemagne, au début des années quatre-vingt, la moitié environ des candidats aux études de médecine étaient favorables à ce critère tout en se montrant très sceptiques quant à l'utilisation de ce seul critère (Trost, 1985).

#### **4.2. Tests de connaissances**

Le degré d'objectivité des tests de connaissances standardisés portant sur des matières scolaires, tests utilisés dans le cadre de l'admission à l'université, est en général élevé. Le test étant livré dans des conditions uniformes, l'objectivité est réelle dans l'exécution du test comme dans son évaluation, dès lors que les devoirs sont conçus sous la forme de questions à réponses multiples (Multiple-Choice). Le degré d'objectivité est plus faible pour les tests de connaissances à réponses ouvertes. Il en va de même pour la fiabilité.

Selon les résultats concordants de la plupart des contrôles de qualité, la capacité prédictive des tests de connaissances demeure inférieure à celle des performances scolaires du degré secondaire; elle est cependant encore satisfaisante (Alciati, Bonetto & Curtoni, 1993). La valeur pronostique d'un test de connaissances augmente en général avec la similitude des contenus du test et des études ultérieures.

La qualité du pronostic de succès dans les études peut être améliorée par la combinaison des deux critères, notes de fin d'études et résultats obtenus au test de connaissances.

Le coût et le travail nécessités pour le développement et l'expérimentation de tests de connaissances standardisés sont relativement importants. Mais une fois ces tests mis au point et standardisés, l'exécution même par un grand nombre de participants et l'évaluation mécanique se font à bien moindres frais et moindres efforts que, par exemple, l'organisation d'interviews.

On ne peut guère poser d'affirmations générales sur l'acceptabilité des tests de connaissances dans une perspective supranationale.

Les tests de connaissances peuvent exercer des effets indésirables sur les contenus et la structure de l'enseignement. Dans bien des pays, en effet, la qualité d'une école se juge face aux résultats obtenus dans les tests; les écoles s'efforcent-elles aussi de préparer leurs élèves le mieux possible à ces tests. Si les contenus de l'enseignement sont axés en premier lieu sur la teneur des tests et que l'accent est porté sur des aptitudes purement répétitives, c'est au détriment de l'acquisition d'un savoir plus étendu et d'une compréhension approfondie. De plus, les tests de connaissances risquent de faire éclore une "industrie de la formation" si l'enseignement scolaire ne dispense pas, ou insuffisamment, la matière examinée dans ces tests. A noter enfin qu'il est relativement aisé de s'entraîner à court terme aux prestations exigées dans les tests de connaissances, même en recourant par exemple à des cours d'entraînement commerciaux.

### **4.3. Tests de capacité aux études**

Les tests de capacité aux études conçus sous forme de questions à réponses multiples (comme le test suédois ou le test allemand d'entrée à l'université) se distinguent par un très haut degré d'objectivité dans l'exécution du test ainsi que dans l'évaluation des résultats. Néanmoins le test finlandais d'aptitude aux études, qui ne contient que des questions à réponses libres, a obtenu une concordance de jugement à 95 pour cent grâce à la standardisation des schémas d'évaluation et à la formation au travail d'évaluation (Lindblom-Ylänne & Lonka, 1995).

On a enregistré aussi des indices de fiabilité élevés, correspondant aux normes internationales usuelles, pour les résultats du test allemand concernant les études de médecine (Deidesheimer Kreis, 1996; Trost, 1995a).

La validité pronostique des tests d'aptitude aux études est quelque peu inférieure à celle des notes scolaires, mais presque égale à celle des tests de connaissances (Beller, 1993; McDonald, 1975; Trost, 1995b; Whitney, 1989).

Dans la plupart des cas, la combinaison des notes scolaires du niveau secondaire avec les résultats de tests d'aptitude aux études accroît plus sensiblement la validité pronostique que la combinaison notes scolaires - résultats de tests de connaissances portant sur des matières scolaires.

Sous l'aspect économique, il y a similitude entre l'utilisation de tests d'aptitude aux études et de tests de connaissances. Le coût du développement et de l'essai des procédures de test standardisées est élevé; la passation de grands nombres de participants est relativement peu coûteuse et l'évaluation mécanique des résultats est bon marché. Dans la comparaison coût/utilité, il ne faut jamais

oublier le bénéfique potentiel d'une sélection judicieuse permettant de réduire le nombre des erreurs d'orientation, des abandons ou des étudiants perpétuels.

L'acceptabilité du test allemand pour les études de médecine a été étudiée. Il en ressort que plus de la moitié des candidats voient dans ce test une amélioration de la procédure d'admission (Trost, 1993a).

#### **4.4. Interviews**

L'objectivité de l'interview en tant qu'instrument de sélection se mesure au degré de concordance des jugements de deux interviewers au moins sur les mêmes candidats. A ce sujet on ne dispose pas d'informations à l'échelle européenne permettant de juger de la qualité de l'interview en tant qu'instrument d'admission aux études de médecine. Edwards, Johnson et Molidor (1990) qui ont consulté des études américaines ont trouvé des valeurs indicatives pour la concordance de jugement qui s'étendent sur une large période mais se situent surtout en zone basse. La recherche générale sur les interviews montre que l'objectivité de celles-ci peut être considérablement augmentée, à condition que les signes d'aptitude à déterminer soient clairement définis, que des règles précises soient données pour l'évaluation des entretiens et - ce qui n'est pas la moindre des conditions - que les interviewers soient dûment formés à cette activité (Trost, 1996).

La validité pronostique de résultats d'interviews est beaucoup plus faible que celle des critères de sélection discutés jusqu'ici (Edward, Johnson & Molidor, 1990; Trost, 1996); cette infériorité est due en partie à une objectivité presque toujours insuffisante de l'interview.

Même jointe à d'autres procédures diagnostiques, l'interview n'apporte qu'un faible gain de validité prédictive.

D'après certaines indications toutefois (p. ex. Hall, Regan-Smith & Tivnan, 1992), la capacité prédictive de l'interview quant au succès dans les études de médecine serait meilleure pour la partie clinique de la formation médicale que pour les études précliniques.

De même, en ce qui concerne la capacité prédictive de l'interview, des améliorations sont possibles par les mesures décrites plus haut, si l'on se limite à juger des signes d'aptitude qui ressortent bien, voire mieux dans l'interview que dans d'autres procédures d'analyse; par exemple, les aptitudes à l'interaction sociale et à la communication, qui peuvent être observées directement en situation d'entretien.

L'interview nécessite un grand déploiement de moyens, en personnel surtout; il est donc peu économique, tant sous l'aspect technique de la procédure que sous

celui de l'organisation. C'est pourquoi, lorsqu'il est utilisée comme critère d'accès à l'université, on le réserve le plus souvent à un petit nombre de candidats qui ont déjà été examinés selon d'autres critères.

Mise à part l'insuffisance de ses qualités psychométriques, l'interview jouit d'un degré d'acceptabilité relativement élevé, tout au moins auprès des candidats. 77 pour cent des candidats allemands aux études de médecine estiment que l'intégration d'un entretien de sélection améliore la procédure d'admission (Trost, 1993a).

## 5. Résumé

Un regard sur la pratique de l'admission aux études de médecine dans les pays européens montre la grande diversité des critères et des instruments de sélection utilisés: 12 genres de critères d'admission auxquels on fait appel en diverses combinaisons.

La combinaison la plus fréquente est la prise en compte des notes de fin de scolarité et des résultats d'un test de connaissances fondé sur des matières scolaires; suivent, dans l'ordre de fréquence, la combinaison des notes scolaires avec la fréquentation de cours déterminés au degré secondaire, puis la combinaison des notes scolaires et d'interviews et enfin la combinaison des notes scolaires et des tests d'aptitude aux études.

De la littérature traitant de l'évaluation des critères et instruments de sélection, pris séparément ou combinés, on peut en déduire ceci: si l'on considère individuellement ces facteurs de sélection, c'est la prestation scolaire du degré secondaire qui, en comparaison, présente le degré de validité pronostique le plus élevé quant au succès dans les études. Les tests d'aptitude aux études et les tests de connaissances portant sur des matières scolaires se classent au deuxième rang. La combinaison de tests de ce genre avec les performances scolaires livre, en ce qui concerne le succès dans les études, un pronostic de bien meilleure qualité que celui résultant de l'utilisation d'un seul de ces facteurs. Tel qu'il est appliqué le plus souvent pour l'admission, l'interview n'a qu'une capacité prédictive relativement faible en ce qui concerne le succès dans les études. Il existe toutefois des indications sur les moyens d'améliorer l'objectivité ainsi que la validité pronostique de l'interview.

## Bibliographie

- Alciati, A., Bonetto, A. & Curtioni, S. (1993). La previsione del successo degli studenti in medicina et chirurgia. *La Formazione del Medico*, 8, 32-39.
- Baron-Boldt, J. (1989). *Die Validität von Schulabschlußnoten für die Prognose von Ausbildungs- und Studienerfolg*. Frankfurt/M.: Peter Lang.



- Beller, M. (1993). Zulassungsverfahren an israelischen Universitäten: Psychometrische und soziale Betrachtungen. In G. Trost, K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 10. Jahrbuch der Pädagogischen Diagnostik* (S. 115-142). Weinheim: Beltz.
- Deidesheimer Kreis (1996). Hochschulzulassung und Studieneignungstests. Studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus-clausus- und andere Studiengänge. Göttingen: Vandenhoeck & Ruprecht.
- Ebach, J. & Trost, G. (1996). Admission to Medical Schools in Europe. Overview on admission procedures - Evaluation of selection instruments - Samples of assessment elements. Lengerich/Berlin/Düsseldorf/Riga/Wien/Zagreb: Pabst Science Publishers.
- Edwards, J.C., Johnson, M.S. & Molidor, J.B. (1990). The interview in the admission process. *Academic Medicine*, 65, 167-177.
- Hall, F. R., Regan-Smith, M. & Tivnan, T. (1992). Relations of medical students' admission interview scores to their dean's letter ratings. *Academic Medicine*, 67, 842-845.
- Ingenkamp, K. (1989). Diagnostik in der Schule. Beiträge zu Schlüsselfragen der Schülerbeurteilung. Weinheim: Beltz.
- Lindblom-Ylänne, S. & Lonka, K. (1995). Selecting students for Medical School: What predicts success? A cognitive approach. University of Helsinki, Department of Psychology. Unpublished manuscript.
- McDonnell, W. (1975). Testing for student selection at tertiary level: A literature review. Victoria: Australian Council for Educational Research in association with Tertiary Education Entrance Project.
- Trost, G. (1985). Pädagogische Diagnostik beim Hochschulzugang, dargestellt am Beispiel der Zulassung zu den medizinischen Studiengängen. In R. S. Jäger, R. Horn & K. Ingenkamp (Hrsg.), *Tests und Trends 4. Jahrbuch der Pädagogischen Diagnostik* (S. 41-81). Weinheim: Beltz.
- Trost, G. (1989). A nationwide testing program for admission to medical schools in West Germany. In R.C. King & J.K. Collins (Eds.), *Social applications and issues in psychology* (pp. 131-137). Amsterdam: Elsevier Science.
- Trost, G. (1993). Attitudes and reactions of West German students with respect to scholastic aptitude tests in selection and counselling programs. In B. Nevo & R. S. Jäger (Eds.), *Educational and Psychological Testing: The Test Taker's Outlook* (pp. 177-200). Göttingen: Hogrefe.
- Trost, G. (Hrsg.). (1995a). Tests für medizinische Studiengänge (TMS): Studien zur Evaluation. 19. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Trost, G. (1995b). Principles and practice in selection for admission to higher education. In Kellaghan, T. (Ed.), *Admission to Higher Education: Issues and Practice* (pp. 7-15). Princeton, NJ: International Association for Educational Assessment.
- Trost, G. (1996). Interview. In K. Pawlik (Hrsg.), *Grundlagen und Methoden der Differentiellen Psychologie. Enzyklopädie der Psychologie* (S. 464-505). Göttingen: Hogrefe.
- Trost, G., Klieme, E. & Nauels, H.-U. (1996). The relationship between different criteria for admission to Medical School and student success. In D. Ajar (Ed.), *New horizons in learning assessment. Ethical, strategic, methodological aspects*. London, New York: Pergamon Press (in print).
- Weingardt, E. (1989). Untersuchungen über Korrelationen zwischen Reifeprüfungsnoten und Erfolg auf der Universität. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung* (S. 252-255). Weinheim: Beltz.
- Whitney, D. R. (1989). Educational admissions and placement. In R.L. Linn (Ed.), *Educational measurement* (pp. 515-525). New York: American Council on Education & Macmillan Publishing Company.

(auf dieser Seite Tabelle 1 einfügen)

(auf dieser Seite Tabelle 2 einfügen)

## **The Swedish Scholastic Aptitude. Test Research and main findings**

### **I. Wedman and W. Henriksson**

#### **1. Introduction**

Today about 130,000 applicants for studies at the tertiary level sit the Swedish Scholastic Aptitude Test (SweSAT) each year. The test has almost become an "industry" in itself when it comes to selection to higher education. The regular staff responsible for the development of the test, or rather, the test battery, consists of 15 persons (in-house staff), but in total about 100 persons in different positions are involved in the development process. The test has become a regular part of the process of selection to higher education.

In this paper we will present the background of the test battery, some reflections and experiences obtained during the development process, basic issues that have been addressed and met during the existence of the test, and some remarks as to what should be considered when implementing a test like the SweSAT.

#### **The SweSAT Programme**

During the last three decade, the system of higher education in Sweden has been the subject of several commissions and committees as well as part of the general debate about education. Many changes have also been introduced during this period. Part of the interest in and the debate about the higher education system has dealt specifically with the selection problem, that is, how admission to higher education is organised and what rules and instruments characterise the selection system.

During the same period a Swedish admission testing program has been developed - the Swedish Scholastic Aptitude Testing Program, SweSAT for short. It has been used for selection purposes since 1977. Before that time, selection was solely based on the average applicant's mark from upper secondary school, provided that formal requirements were fulfilled for the study program in question. Such formal requirements were expressed as a minimum level based on marks or grades obtained in various subjects.

In 1977, the system was changed. For applicants who were older than 25 years and had more than four years of work experience, the SweSAT provided an opportunity to gain access to higher education in case there were more applicants than places and selection was necessary (see also Wedman, 1994).

#### **Some background information**

During the 1960s the higher education system began to change in Sweden. Since then several government commissions have discussed and proposed further changes of the system. A special aspect of interest has been the admission system, that is, the selection of applicants for many of the various study programs. The latest commission report was presented in 1985 and led to a major change of the system, including a great expansion of SweSAT use from 1991 onward.

A parallel and very important development has concerned the marking/grading system used in primary and secondary schools. From the early 1960s onwards the marking system in Sweden has been norm-referenced, that is grading on a curve. In 1995, this was changed to a criterion-referenced system. This change will eventually affect the selection system used when admitting students to universities and colleges. Given these circumstances, it is expected that the use of the SweSAT programme will expand further.

### **The SweSAT Programme**

As part of the change in the higher education system in the 1960s, the admission system was reviewed. An expert group led by Professor Sten Henrysson began investigating the possibility of developing an entrance test to be used for all applicants to higher education. Many of the various tests applied in other countries were studied and put to trial use.

There were several arguments put forward for developing an entrance examination of the type that the SweSAT represents. One of the main arguments was the need for means by which to compare applicants who lacked comparable marks. Many of those applicants belonged to the so-called 25:4 group, that is, applicants who were at least 25 years old and had at least four years of work experience. There was a politically strong and well-articulated ambition to make colleges and universities accessible to more people.

Another argument for developing an entrance test was the idea that students from upper secondary school should be given a second chance to gain access to colleges and universities. At the time, only marks were used when selecting students for entrance to colleges and universities. A second route was sought to reduce the emphasis placed on marks or grades.

In 1975, the Swedish Parliament decided to change the admission system beginning with the autumn 1977 student selection. However, the decision provided that the use of the entrance test should be limited to applicants who were at least 25 years old and had at least four years of work experience. The argument regarding a second chance for younger students was to be reconsidered when the new system had been in operation for some time.

The decision to limit the use of the SweSAT program led to about 10,000 applicants taking the test each year from 1977 to 1989, about 6,000 at the spring administration of the SweSAT and 4,000 at the autumn testing.

As a result of the government commission report about changing the admission system was presented in 1985, the use of the SweSAT programme has increased dramatically. The commission suggested that the SweSAT programme should be open to all applicants, including those leaving upper secondary school. Beginning with the 1991 autumn administration of the SweSAT, selection to higher education is based either on marks from upper secondary school or on SweSAT results. So far, about per cent of the available places in study programs with a limited number of places have had applicants selected on the basis of average marks and about 40 per cent on the basis of SweSAT result.

Applicants are judged on the most favourable condition. They do not have to decide themselves whether to compete on the basis of their average mark or their SweSAT result, which has also meant that most applicants choose to take the test. About 130,000 students do so each year.

## **2. The present content of the SweSAT**

The SweSAT has been used in selection to higher education since 1977. At first the SweSAT was open only to applicants who were at least 25 years old and who had at least four years of work experience. As from 1991 the SweSAT may be used by all applicants to higher education. The selection procedure is based on three groups of applicants; those with GPA (grade point average) from upper secondary school, those with SweSAT-score, and those with SweSAT-scores with additional scores for work experience (WE). Applicants can belong to more than one group and, if so, the student is selected in the group in which he/she has the best rank. About 60 per cent of the places available at a certain study programme go to applicants selected on the basis of GPA and about 40 per cent to applicants selected on the basis of test score.

The content of the SweSAT remained almost the same until 1996. Until then, the SweSAT consisted of six subtests and the items in a certain version (two versions are administered each year) were public documents as soon as the sixth and last subtest of a SweSAT administration had been completed by the test taker. Some minor changes of the SweSAT programme were conducted between 1977 and 1996. Worth mentioning is that in 1992, ERC, a reading comprehension test in English, was introduced. ERC replaced the subtest aiming at capturing the complex skills of study techniques.

Since 1996 the SweSAT consists of five subtests, as the general information subtest has been eliminated. The SweSAT is now administered in five blocks of

which one block consists of try-out items. The items in the try-out block are not public. The total test time is 50 minutes for each block, two blocks consist of the same subtests. The design of the administration is illustrated in Table 1 and the SweSAT programme is described in Table 2.

*Table 1: Regular test items and try-out items in the SweSAT programme divided into blocks.*

| <b>Block</b> | <b>Subtest</b>                              | <b>Abbreviation</b> | <b>Time</b> |
|--------------|---|---------------------|-------------|
| 1            | Interpretation of diagrams, tables and maps | DTM                 | 50          |
| 2            | English reading comprehension               | ELF                 | 35+         |
|              | Vocabulary                                  | WORD                | 15          |
| 3            | Reading comprehension                       | READ                | 50          |
| 4            | Data sufficiency                            | DS                  | 50          |
| 5            | Any block of above                          |                     | 50          |

The order of the blocks (1-4) varies from one administration to another. The items in the fifth block are try-out items. These are not included in the test takers score.

Table 2: The Swedish Scholastic Aptitude Test (SweSAT).

| Subtest                                     | Abbreviation | Number of items | Time         |
|---|--------------|-----------------|--------------|
| Interpretation of diagrams, tables and maps | DTM          | 20              | 50           |
| English reading comprehension               | ELF          | 20              | 35           |
| Vocabulary                                  | WORD         | 40              | 15           |
| Reading comprehension                       | READ         | 20              | 50           |
| Data sufficiency                            | DS           | 22              | 50           |
| Total                                       |              | 122             | 3 hrs 20 min |

### 3. Five major problems encountered

During the development phase of the SweSAT Program we have encountered five major problem areas which may be categorised as follows:

1. Social bias
2. Sex bias
3. Coaching
4. Prediction of academic performance
5. Practical matters

Each of these problem areas will be briefly described below.

#### **3.1. Social bias**

During the development phase, the issue of social bias in a test like the SweSAT Programme was heavily discussed, and primarily from a political point of view. It should be remembered that the test was introduced primarily to make higher education accessible to persons that earlier had been shut out from studies at the tertiary level. It is also important to note that Sweden this time had a social



democratic government. The very reason for introducing the SweSAT Programme was to find a way to improve opportunities for people that, generally speaking, represented lower socio-economic groups.

If one divides all applicants into socio-economic groups, all statistics show that those from the upper groups obtain a higher average score on a test like the SweSAT than those from lower socio-economic groups. The same picture will appear for almost every cognitive measure known of, including grades. The reason for this is quite simple and has primarily nothing to do with socio-economic status per se, but rather with the fact that applicants from higher socio-economic groups enjoy better study conditions than applicants from lower socio-economic groups.

To some influential Swedish politicians, this came as a bite of a surprise. The future of the test was up for discussion. However, there was, as it seemed, no third way out of the dilemma of introducing an admission test for older applicants (without comparable merits) and at the same time having a part of so called social bias in such a test. The discussion about social bias ended more or less and the idea of introducing the test gained new strength. When the test was introduced in 1977 it was supported by almost all interest groups.

In retrospect, social bias was not the issue of greatest concern when the test was introduced in 1977. Instead, the idea of measuring cognitive achievement by a test that took about four hours was. At this time, there was a very strong movement in Sweden for abolishing any written instrument that proposed to measure cognitive achievement. Students and teachers all over Sweden, primarily at the university level, questioned every form of written assessment and suggested instead group work and group assessment in its widest sense, a movement that ended in the beginning of the 80s.

The scientists that had developed the test were greatly concerned when it was introduced and feared massive criticism, given the circumstances mentioned above. However, nothing happened. No criticism, was voiced. The explanation for this has remained somewhat unclear, but perhaps the founder of the test, Professor emeritus Sten Henrysson, was right when he pointed to a number of significant events took place around the world on the very day that the test was introduced, and that these events caught most of the mass media interest.

### **3.2. Sex differences**

In the 80's, the debate on social bias was replaced by the debate on gender. In fact the gender issue has influenced the work with the SweSAT Programme to a much greater extent than the social bias issue. The basic reason for this is that males generally speaking, outperform females on tests like the SweSAT. In the western part of the world, thousands of studies have been conducted in order to

find underlying causes. We have ourselves conducted a number of studies in this field. There seems to be no clear-cut explanation for this fact, even though a number of hypotheses have been presented.

The difference is rather small but systematic. It is important to bear in mind that the difference appears on a group level. In fact the distributions of males and females overlap to a very great extent, which means that a huge number of females outperform males on the SweSAT. Yet, on a group level, males outperform females.

One hypothesis presented is that the very format of the SweSAT and similar test batteries works in favour of males, i.e. males seem to handle the multiple-choice format better than females do. This is, however, a question under debate. On the other hand, we know that females usually outperform males on essays. Results from a number of studies indicate that this is the case, although the reasons remain unclear. Therefore, to incorporate an essay part in the SweSAT Program would likely lead to a test battery in which the differences in achievement between males and females would vanish. What makes the idea of incorporating an essay part unrealisable within the SweSAT Programme and most other programmes of this kind, are the costs involved. In order to obtain reliable results in scoring an essay you need at least three independent scorers, which in our case means 390,000 scorers per year!

The issue of gender differences should, according to our experiences, be approached with great seriousness. Besides investing a great amount of money in doing our own research in the field (Stage 1993), we have taken measures to make sure that men and women are equally represented in different expert panels that examine the test before it is presented. Since a few years back the government group for equal treatment (of males and females) is free to send a member to participate in the development and the final examination of a new test version.

### **3.3. Coaching**

Henriksson (1981) made a review of relevant studies of the effects of practice and coaching on test score and his conclusion was that the effects caused by low test experience and a complex item format are the two main factors leading to score gains.

Two different approaches to the issue of special preparation and coaching. The first approach is to ensure that all test takers are familiar with the test and with the strategies required to take it. This approach will be discussed below under the heading "The SweSAT and Test-Wiseness".

The second approach is to develop tests free from items susceptible to special training, cramming, and coaching. This latter approach will be discussed in the next section in connection with Data Sufficiency items.

### **Data Sufficiency (DS) and Item Format**

It has often been claimed that items with a complex format are susceptible to coaching (Vernon, 1960; Pike & Evans, 1972; Powers, 1986; Becker, 1990):

The results of this review support the hypothesis that complex item formats are indeed more susceptible to coaching, practice, and/or test preparation than are simpler formats. The number of words used to convey directions seems to be a significant aspect of this complexity, as does the fixed format nature of many of the most susceptible item types (Powers, 1986, p 76).

However, increasing an examinee's familiarity with novel item types (such as the SAT-M's datasufficiency or quantitative-comparison items) may well enable him or her to improve SAT performance considerably (Becker, 1990, p. 404).

In a series of studies Henriksson (1981) investigated DS items and their susceptibility to special instruction and coaching. The basic idea for the coaching intervention was that a rational problem-solving behaviour for DS items could be modelled as a sequential process including 2 (or 3) subproblems, i.e. 2 (or 3) decisions (Yes/No) in relation to each subproblem. The DS item format was also reduced in complexity by this transformation of the original problem to 2 (or 3) simpler subproblems which were to be answered separately. The coaching also included information about how to use partial knowledge.

A brief summary of the results of these three studies is that the test score was affected neither by coaching nor by practice. These conclusions also applied to conventional scoring (0-1) as well as to differential scoring. The purpose of the differential scoring model was to reflect partial knowledge.

In three studies (Henriksson, 1981) a special item format was developed with the instructed strategy as a main point of departure. This item format made the test taker act automatically according to the sequential process of problem-solving which was instructed orally in the earlier studies. The results of these three studies indicated that the obtained score was not affected by item format; neither if the score was based on conventional scoring nor on differential scoring.

To sum up, the conclusion drawn from the studies by Henriksson was that the sequential problem-solving strategy, which was instructed as well as induced in the item format in order to obtain a logical and rational problem-solving procedure, had no effect. Regarding the question of the complexity of the DS item format the results also indicated that the DS format is not complex, at least not

for students from grade three in upper secondary school. This means that the DS item format optimises the relation between actual knowledge and obtained score for a test taker, i.e. the score reflects the test taker's, capacity in terms of full as well as partial knowledge.

### **The SweSAT and Test-Wiseness**

If test takers have little or no experience of tests and test taking, their obtained scores may be too low and not in accordance with their actual capacity. A fundamental assumption for test-fairness in this respect is that the test takers are test-wise. Consequently, the required test-wiseness should be made available to all potential test takers. With reference to the SweSAT, there are two different approaches to the problem of test-wiseness.

Everyone who registers for the SweSAT gets the information brochure "Högskoleprovet". The aim of this material is to familiarise the test takers with the SweSAT and to inform them about different circumstances and details before, during and after the testing. The material also presents the item format and sample items for each of the subtests, and the potential test taker is also requested to take this mini-SweSAT within the given time.

Another circumstance, contributing to test-wiseness and therefore worth mentioning, is that four out of five administered blocks of SweSAT-items are public documents. This means that all items, except items in the try-out block, are public. Consequently a potential test taker can get access to, for example, the latest SweSAT version and, thus, may practise on a full sample test.

The summarised judgement is that access to older versions of the SweSAT, together with the familiarisation material, create optimal conditions for test-fairness for the SweSAT with respect to test-wiseness.

### **3.4. Prediction of academic success**

Attempts to predict academic success have gained much attention and several studies have been carried out to get information about the design of the admission procedure. Most universities utilise selection procedures based on individuals, previous performance, such as college grades (GPA) or results on tests, such as SAT, ACT or SweSAT, which are designed to predict general academic success.

Predictive studies are important for the evaluation of the strength of an admissions test like the SweSAT. As a consequence, thousands of predictive validity studies have been carried out over the years within the field of educational measurement, e.g. Willingham et al (1990); Henriksson & Wedman (1992). The majority of these studies have been reported as a single correlation coefficient

with the grades obtained after one or a few years of academic studies. However, the correlations and the results of predictive studies are often very difficult to interpret because of a number of statistical and other problems. Only admitting those who are at the top of the ability scale leads, for example, to the well-known restriction of range problem. The reliability of the criterion as well as a very restricted criterion scale, which is the case with using pass/fail grading scales, are other problems with predictive studies -(Henriksson & Wedman, 1992).

### **SweSAT and predictive studies**

In order to get information about the predictive validity of the SweSAT, a rather elaborate predictive study was carried out in the seventies. This study was based on a predecessor to the SweSAT, e.g. Henrysson & Wedman (1979). The subjects (n=584) were students of engineering at two technological institutes and students at teacher training colleges. Test score, admission points (GPA) and a relatively refined criterion scale, based on a summary of a number of exams, were collected. Information was also collected on, among other things, study results, change of schools, and study interruptions. After the first year information was obtained about the points acquired on each subcourse. After the second year the same information was collected again. This concluded the data collection for the technical colleges, while information was once more - three years later - obtained from the teacher training colleges. The results obtained showed a fairly high correlation (about 0.4-0.5) between the test and the criterion, which was interpreted as a support for the hypothesis that SweSAT could compete with grades in predicting academic performance.

Henrysson et al (1985) carried out a validity study which consisted of 12 study programmes at four universities (n=3 171). The main result of the study was that students with three years of upper secondary school was the best performing group. Academic success was defined as the total number of credits achieved by the students. On the whole, GPA and SweSAT showed the expected correlation with academic success, i.e in the interval 0.35-0.45. Another result was that work experience (WE) seemed to reduce the predictive validity of the admission instrument (GPA or SweSAT-score). Three explanations were put forward. (1) Students admitted by way of WE often had a lower GPA than students without WE. (2) Students admitted by way of WE were older than other students and lacked previous experience of studies. (3) The older, WE-admitted students in general already had families and must provide for these. Such conditions could be thought to give WE-admitted students less time to concentrate on their studies.

In the early 90s new questions were raised about the predictive validity of the SweSAT, basically because its use in the selection procedure had been expanded

to allow all students to take the test. These changes created a need for a validity study of current interest. The results of validity studies are usually expressed as a correlation between the current predictor and some form of criterion of academic success. Since there are recurrent and difficult methodological problems associated with this type of design, i.e. Henriksson & Wedman (1992), a study was carried out with a different design. A sample of students from three different admission groups was followed during three years of academic studies. The study will be published in *Scandinavian Journal of Education Research* in 1997 (Henriksson & Wolming). In the following section we will highlight the main findings of this study.

In this study student performance during a period of three years was observed for a sample of 840 students from four different study programmes (Business administration, Medicine, Technical engineering physics, and Social works). All students at these study programmes at two out of five universities were selected. The students were divided into three groups, according to the grounds on which they had been admitted to the programmes. The first group of students had been admitted on the basis of earlier academic achievement (GPA), the second group on the basis of their SweSAT-scores, and the third group on the basis of their SweSAT-scores with additional scores for work experience (WE). The purpose of the study was to see if there were any differences in academic performance at the programmes between the three groups. Academic performance was defined in three ways:

- a) the number of credits achieved after each semester
- b) drop-out frequency after each semester
- c) the number of credits achieved after each semester by those students who completed their studies.

Swedish universities and colleges use a system of credit points for studies on the undergraduate and graduate levels. One point corresponds to one week of full-time studies. One academic term comprises 20 weeks (20 points). 40 credits represent one full academic year. The number of credits acquired by all students was registered at the end of each semester. Information about study results, predictor information (GPA, SweSAT, WE), and the grounds on which a certain student had been admitted to the programme was obtained from the national student records database.

\* Credits achieved after each semester

The first criterion variable in the study was the number of credits acquired over time by the three admission groups. The drop-out frequency was not taken into account in this first analyse.

Figure 1 shows the performance of the three admission groups after each completed semester.

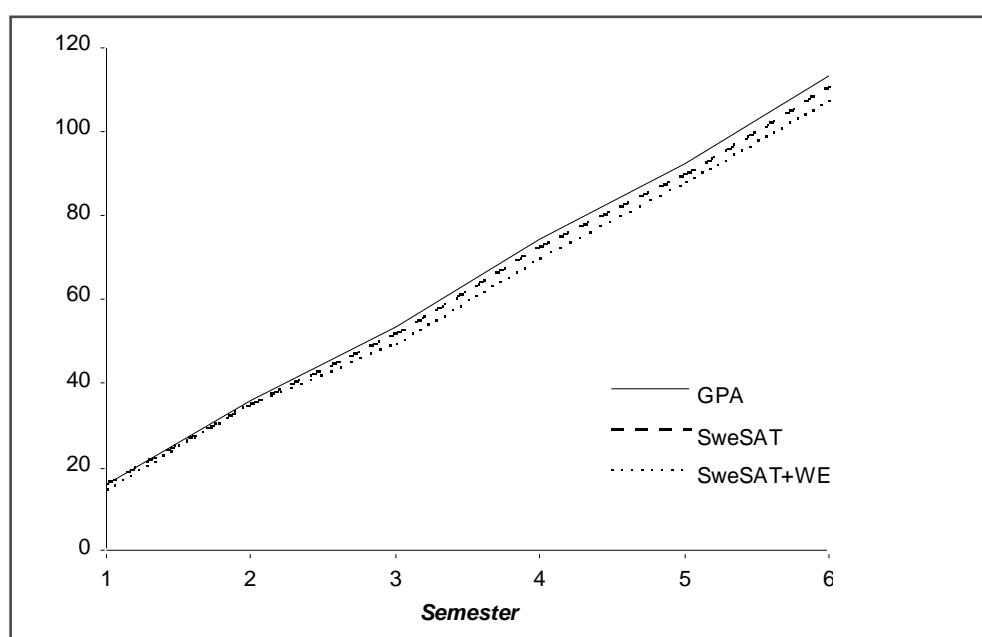


Figure 1. Number of credits acquired after each semester by the three admission groups. All study programmes.

At first glance, the differences between the admission groups as displayed in Figure 1 seem relatively small. However, the statistical analysis (Kruskal-Wallis 1-Way Anova) indicated a significant difference between the admission groups ( $p < 0.05$ ).

A pairwise follow-up analysis with the Mann-Whitneys U-test on all study programmes indicated that the differences were to be found between the students admitted on the basis of GPA and the students admitted on the basis of SweSAT+WE. The differences between the two groups appeared after semester 2, 3, 4, 5, and 6. Significant differences were also found between the two groups admitted on the basis of SweSAT (with and without WE). The differences between these two groups appeared after semester 2, 4, and 6. No significant differences were found between the students admitted on the basis of GPA and on the basis of SweSAT.

The analyses above were also made for each of the four study programmes. The significant differences which appeared between the GPA-group and SweSAT+-WE-group in the total group (all study programmes) could be related to the Business administration and economics study programme. At the other study programmes, no differences were found between the admission groups.

\* Drop-out frequency after each semester

The second criterion variable was the drop-out frequency after each semester in the three admission groups. This is shown in Figure 2.

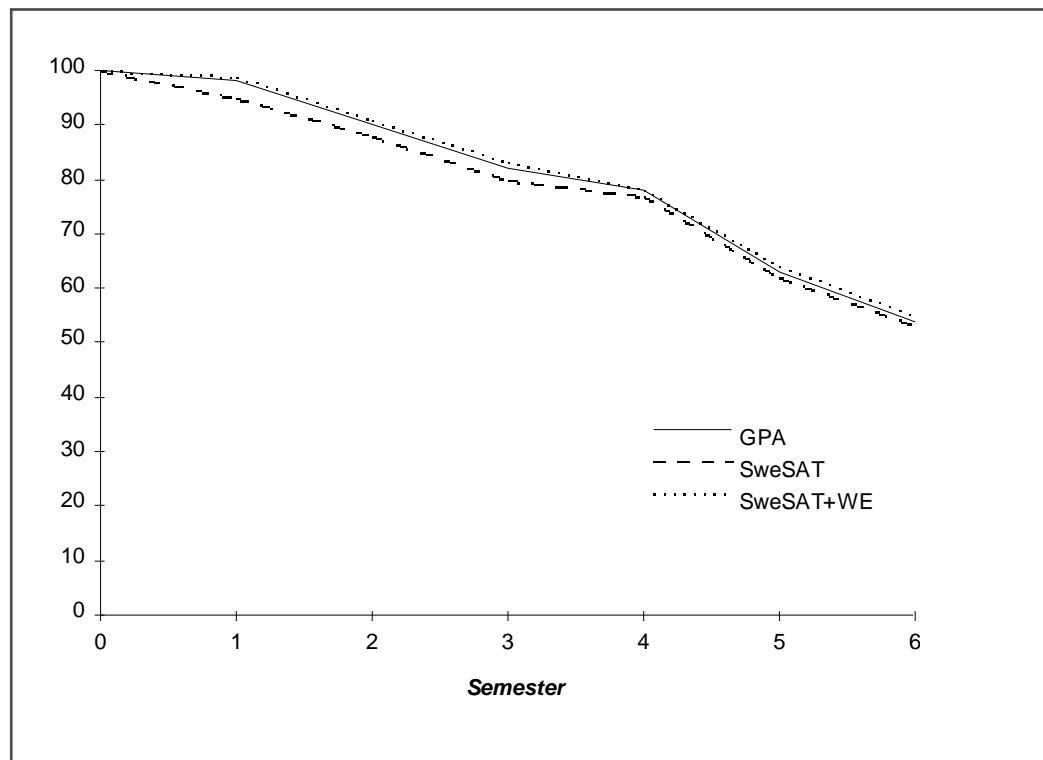


Figure 2. Drop-out frequency after each semester by the three admission groups. All study programmes.

As shown in Figure 2, the drop-out frequency is very much the same in the three groups. The Chi-Square analysis did not indicate any significant differences between the groups in this respect. When the study programmes were tested one by one, differences appeared at the Medical study programme. As shown in Figure 3, the drop-out frequency is higher in the WE-group than in the other two groups, and the dropping out takes place after the second semester. The figure also shows that dropping out takes place in the SweSAT-group after the third semester.



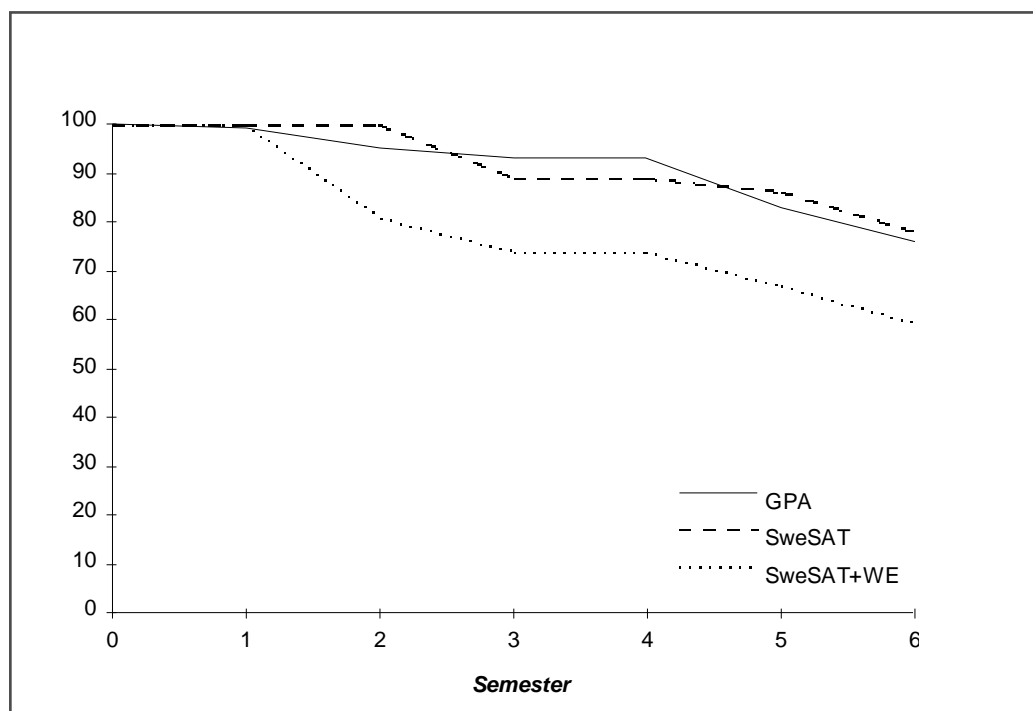


Figure 3. Drop-out frequency after each semester by the three admission groups. Medical study programme.

The drop-outs in the WE-group after the second semester consisted of five students. The drop-outs which were found in the SweSAT-group after the third semester consisted of three students. It can also be noted that after the second semester, the inclination of the drop-out curve was about the same for each of the three admission groups (Figure 3).

\* Credits after each semester for those students who completed their studies

The third and final criterion variable used in this study was the number of credits acquired after each semester by the remaining students in the three admission groups. The main reason for making a distinction between success for the total group and success for the remaining group was that not considering drop-out frequency might give an incorrect picture of success if there are systematic differences between groups and study programmes.

The average number of credits acquired by the remaining students closely resembles the outcome of the first variable, which did not take the drop-out frequency within the groups into account (Figure 1). The statistical analysis (Friedmans Two-Way Anova) indicated that significant differences ( $p < 0.05$ ) between the admission groups could be found in the total group (all study programmes) and at the Business administration study programme.

The results of the total group are shown in Figure 4, and indicate that the GPA-group shows the highest level of performance when compared to the other

groups. The SweSAT-group is followed by the SweSAT+WE-group which shows the poorest performance. The differences which appeared in the total group were located between the GPA- and the SweSAT+WE-group and occurred after all semesters except the first one.

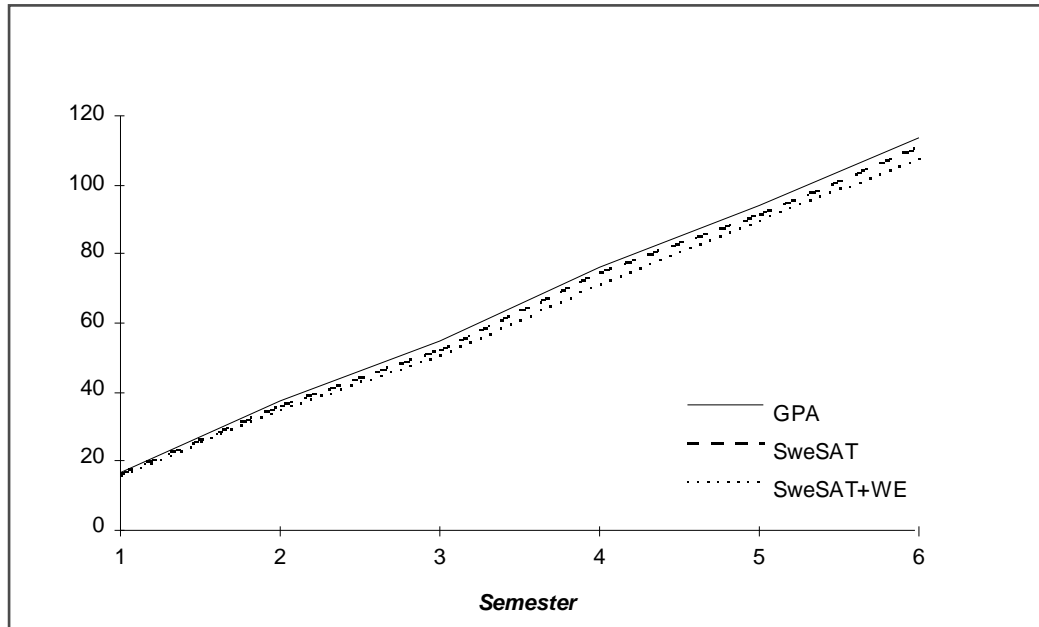


Figure 4. Number of credits acquired after each semester by the three admission groups. Remaining students. All study programmes.

The differences found in the total group could be related to the Business and administration study programme. Friedmans Two-Way Anova showed no significant differences within the study programmes, except for the admission groups at Business administration.

The drop-out frequency (Figure 3) after the second semester among students who had been admitted on the basis of SweSAT+WE at the medical study programme raises questions.

One question is whether students admitted on basis of their SweSAT+WE score were less successful in their studies at the Medical study programme. A closer look at Figure 5 reveals that this is not the case. According to this figure there are hardly any differences at all between students admitted on the basis of GPA, SweSAT and SweSAT+WE, respectively.

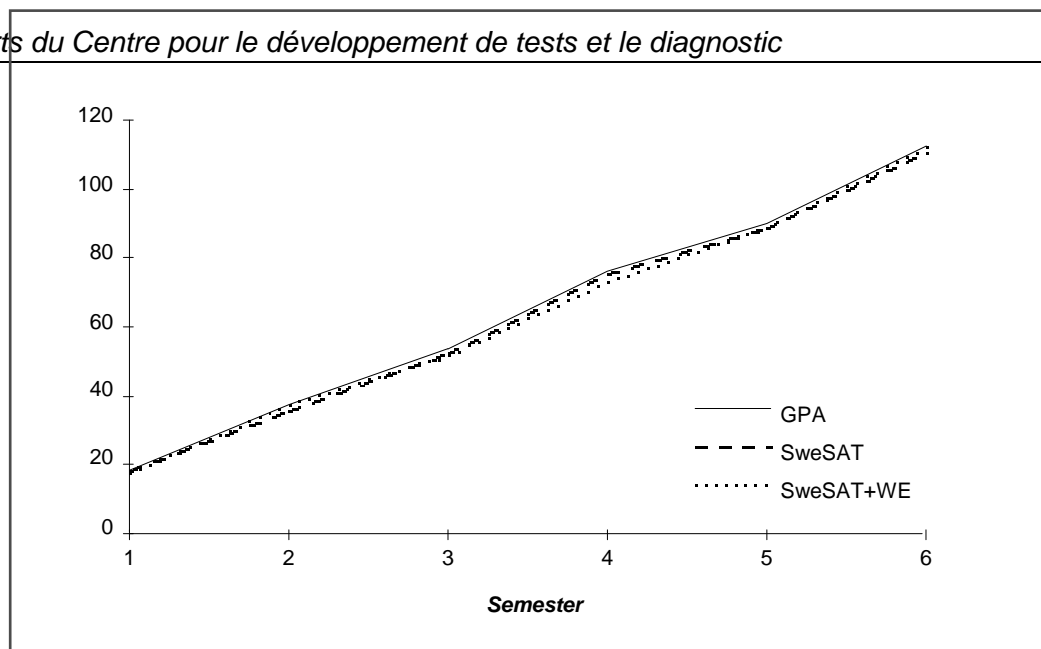


Figure 5. Number of credits acquired after each semester by the three admission groups. Remaining students. Medical study programme.

Another question to consider when analysing drop-out frequency concerns the reason for dropping out. In some cases, this has no relation at all to study success. For example, the cause for dropping out may be that students change their field of study, that they stick to the study programme but move to another city, or that they get an employment.

The summarised and over-all conclusion from this study is that the selection procedure based on three admission groups, which has been used in Sweden since 1991, gives about the same results in terms of study success for students selected on the basis of GPA and on the basis of SweSAT-score. Maybe some questions could be raised about the success for students selected on the basis of SweSAT+WE at the Business administration study programme, but, on the other hand, the differences at the other study programmes were very small. At the medical study programme there were, for example, hardly no differences at all (Figure 5). It can also be added that the reason for giving merits for WE is the notion that students with WE in general give other students and the whole educational situation a broader frame of reference which is supposed to be fruitful in the academic setting.

### **3.5. Practical matters**

Aside from the problems mentioned above, there are some miscellaneous issues that should be taken into account when introducing a test like the SweSAT Program. One of these is the issue of security. In our case this has led to the development of a very strict procedure for printing, distribution, supervising the test administration, and for collecting and scoring the tests. The costs involved for these different measures have been quite large.

When a new version of the SweSAT is ready for printing, two staff members take it to the printing office which in our case is located a few miles from our office. The test version is handed over to the head of the printing office who then brings into a separate and locked room within the printing building. When the test is ready for distribution the official post service collects the tests and place them in sealed boxes for separate distribution to the different test centres all over Sweden (mostly universities). When the tests arrive there, a head test administrator (specifically trained for handling and administering the tests) takes over and locks up the tests.

At the test administration every applicant has to identify himself/herself and a seat is given to him/her by the head test administrator. At the end of the session, every answer sheet is checked against the applicant's ID. The answer sheets are sent to the computer department at our university where all sheets are checked against lists of the applicants participating at the different test centres. The process of scoring (optical scoring) then takes place in a locked room to which only the head of the computing process has access. In our case that means that as being chairmen of the Department of Educational Measurement and having overall responsibility for the test, we do not have access to the "scoring room" without permission from the head computing officer.

After the results have been obtained they are examined. Seemingly aberrant results are checked specifically and, if needed, then reported to the Agency of Higher Education that formally is in charge of the test.

#### **4. Concluding remarks**

Implementation of an admission program using a test battery like the SweSAT involves several components, some of which are more or less of a technical character. But there are also other components to pay attention to, among them political considerations. The story of SweSAT Program tells quite a bit of such political considerations.

Technically we know quite well how to develop an instrument to be used for selection purposes. In this respect we all can rely on a history of 100 years which started its second phase by the work of Alfred Binet and Henry Simon in 1905 (see for example Du Bois, 1970; Eswards, 1971; Linden & Linden, 1968). They proved, that by using a test consisting of 36 items it was possible to rank young pupils concerning their cognitive achievements. Since then remarkable technical progress has been made, but the basis for ranking students is more or less the same. The use of written tests of different kinds is still a very effective way of reflecting cognitive development and achievement.

The SweSAT Programme consists of five subtests of different kinds, ending up in 122 items in total. In principle it would be possible to use much fewer items

and still make the instrument effective for selection purposes. To do so would, however, make the test unacceptable for other reasons than technical ones. First of all it would be politically impossible, at least in Sweden, to do so. Such a test would not match the expectations of a serious instrument that covers a main part of what we call cognitive achievement is aimed at predicting future achievement. Secondly, the backwash effect would be tremendous, as students and their parents would request courses on such specific content as is reflected by the test. Other topics would soon be left out. On a more general level this is already the case, i.e. the content of an admission test will always have an effect on what is taught at study programs leading up to admission to further education. This, we think, is important to bear in mind when decisions about developing an admission test are to be made.

The arguments above quickly lead us on to the three critical elements in test development, namely:

1. The test has to be technically sound fulfilling basic requirements of reliability and validity
2. It has to match basic expectations of those taking the test, people at large and not the least politicians
3. It has to be cost-effective.

These three elements are not easy to combine. Test development is rather a question of balancing them. A test battery like the SweSAT is very much coloured by the efforts to do just that. We think that we have managed to develop a test battery that matches the three elements in a satisfactory manner. Its technical standard is very high (the easiest element to match), it has generally been very well received by the applicants and others interested in the admission process, including politicians (the test is evaluated on a continuous basis), and it is very cost-effective (the total cost per applicant is less than \$25).

Programmes like the SweSAT are faced with a number of more specific issues than the ones just mentioned. The SweSAT Programme has (aside from a number of technical issues) been faced with primarily five issues, namely 1/social bias, 2/gender bias, 3/coaching, 4/prediction of academic performance, and 5/security matters. All these issues are examples of such considerations that have to be dealt with implementing a selection instrument like the SweSAT Programme. The importance of each one of these issues is probably dependent on each country's culture and other unique circumstances, but seen from a western world perspective each one of them has played a major part in the development of many testing programs in many countries.

The issue of social bias has to be treated seriously. The fact that a testing programme like the SweSAT will result in applicants from higher social groups achieving a higher score (on average) than applicants from lower social groups must not be seen as inevitable, i.e. every measure possible should be taken to limit the effect of social bias. Every item included in a selection test should be examined carefully to make sure that its content can be defended as fulfilling the basic aim of the test and not giving special advantages to certain groups of applicants.

The gender issue has been very much debated and researched. The measures mentioned with reference to social bias should be applied also in this aspect. It is a must that each individual item in a test be scrutinised as to the rational why it is "favouring" males (most typically the case) or females.

Coaching seems to be a never-ending story in the business of using tests for selection purposes. Even though the effects, according to the research that has been conducted so far, most often seem to be small we have to live with the fact that applicants might think otherwise. There are (at least) two sides of this coin. One is to make sure that the items and tests used are not coachable. Some item formats seem to be more coachable than others. In the process of developing the SweSAT Programme some 50 prototypes of item formats were tried out. Most of them were left out as they proved to be coachable. The other side of the coin says that it is of uttermost importance that the applicants have a fair chance of getting acquainted with the item formats used. From the results of different coaching studies carried out so far, we know that applicants often improve their performance from the first time the test is taken to the second time and that this improvement most probably has to do with a better acquaintance with the item formats used.

The issue of prediction of academic performance has been with us since long ago in fact since long before Binet and Simon presented their test that meant a new start concerning how to measure cognitive achievement. Thousands of studies have been carried out in this field and tons of research articles have been presented. The results are generally the same in most of these studies. Grades and tests like the SweSAT Programme are the best predictors known for foretelling later achievement. Most of the variance in later achievement that is to be explained will be covered by these two instruments. Still, there is a must for every testing program to continue to conduct its own studies. The issue of prediction of academic performance exemplifies that the development of a testing programme for selection purposes involves a lot more than technical issues.

The security problem in any large scale testing program like the SweSAT Programme must be dealt with in a very cautious way. Every step in the process of

developing the test, printing it, distributing it, administering it, collecting the results and scoring it has to be carefully laid out.

Finally, a few words about the future of selection tests seen from a Swedish perspective. The SweSAT Programme was introduced at a time when there was a very critical atmosphere concerning every kind of individual given written tests. We sort passed that period with honour, i.e. the test seemed to be accepted by most persons, applicants, people at large and politicians. The 80's saw an even greater acceptance of the test. In the late 80's voices were raised concerning for broadening the set of instruments used for selection to higher education, and most colleges and universities were given larger freedom to develop their own instruments parallel to grades and the SweSAT Programme. Many of them took the chance of developing their own specific instruments, mostly interviews, portfolios of different kinds and more content directed tests. It is still too soon to evaluate these efforts, but up until today no results tell the story that a replacement of grades and the SweSAT Programme by these new instruments would guarantee a better prediction of academic performance. However, a two-step procedure might be a way to follow, where grades and results from the SweSAT Programme, constitute the first part with more specific instruments for the training program in focus as complements but again such a step has to be measured against the third element, namely cost-effectiveness.

## References

- Becker, B.J. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research*, 60, 373-417.
- Du Bois, P.H. (1970). *A history of psychological testing*. Boston: Allyn and Bacon, Inc.
- Edwards, A.J. (1971). *Individual mental testing Part I: History and Theories*. San Francisco: Intext Educational Publishers.
- Henriksson, W. (1981). Effekter av övning och instruktion på testprestation. (The effects of practice and coaching on test score). Dissertation, Department of Education, University of Umeå.
- Henriksson, W., & Wedman, I. (1992). Prediction of academic success in a perspective of criterion-related and construct validity. *Educational Measurement*, No 2. University of Umeå, Department of Education, Division of Educational Measurement.
- Henriksson, W., & Wolming, S. (in press). Academic performance at four study programmes. A comparison of students admitted on the basis of GPA and SweSAT-scores with and without credits for work experience. To be published in *Scandinavian Journal of Educational Research*.
- Henrysson, S., & Wedman, I. (1979). Research in Sweden with regards to predictive tests and interviews for admission to higher education. *Spånor från SPINT*, Nr 13. Department of Education, University of Umeå.
- Henrysson, S., Kriström, M., & Lexelius, A. (1985). Meritvärdering och studie-prognos. Några undersökningar av antagningssystemets effekter (Evaluation of merits and academic performance. Studies of the effects of selection regulations). *Arbetsrapporter från*

pedagogiska institutionen, Umeå universitet, Nr 21, Pedagogiska institutionen, Umeå universitet.

- Linden, K.W., & Linden, J.D. (1968). *Modern mental measurement: A historical perspective*. Boston: Houghton Mifflin Co.
- Pike, L.W., & Evans, F.R. (1972). The effects of special instruction for three kinds of mathematics aptitude items. College Entrance Examination Board: Research and Development Report 71-72, No. 7 and ETS Research Bulletin 72-19. Princeton, New Jersey: Educational Testing Service.
- Powers, D.E. (1986). Relations of test item characteristics to test preparation/test practice effects: A quantitative summary. *Psychological Bulletin*, 1986, 100, 67-77.
- Stage, C. (1993). Gender differences on the SweSAT. A review of studies since 1975. *Educational Measurement*, No 7. University of Umeå, Department of Education, Division of Educational Measurement.
- Vernon, P.E. (1960). *Intelligence and attainment*. London: University of London.
- Wedman, I. (1992). Selection to higher education in Sweden. *Educational Measurement*, No 1. University of Umeå, Department of Education, Division of Educational Measurement.
- Wedman, I. (1994). The Swedish Scholastic Aptitude Test: Development, Use, and Research. *Educational Measurement: Issues and Practice* 5-11.
- Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990) *Predicting college grades: An analysis of institutional trends over two decades*. Educational Testing Service, Princeton.



## **The Medical College Admission Test (MCAT) - its use in U.S. and Canada and some results of validation studies**

**J. L. Hackett**

### **Medical Education in the United States - Background Information**

Medical education in the United States is provided by a mix of independent, private medical schools or colleges as well as medical schools supported or operated by the education system of one of the 50 state governments. There is only one Federally operated medical school, and it was established to train physicians for duty with the Armed Forces. There are 125 accredited schools of allopathic medicine in the United States. Successful graduates of these schools earned a Doctor of Medicine degree (M.D.). Other health professions institutions train students to become doctors of osteopathic, podiatric, or veterinary medicine. Most medical schools are university-based, and most require applicants to have successfully completed a course of studies in the sciences and liberal arts at an undergraduate, baccalaureate-degree granting college or university before they can matriculate to medical school. Many medical schools are associated with teaching hospitals at which medical school graduates receive their residency and medical specialty training following award of their M.D. degree.

In 1890, 66 medical school deans, motivated by a common desire to elevate the standards of medical education, met to form the Association of American Medical Colleges (AAMC). Initially, the primary focus of the Association was to aid the implementation of major reforms in academic medicine recommended by national studies, such as university-based medical education. The Association then turned its attention to improving the process of medical education. In the 1960s, the Association reorganized to support the full range of its members concerns -- education, research, and service to patients -- giving teaching hospital executives, medical school faculty, and medical students a voice in its governance. Today, the AAMC is a private, nonprofit association with its membership comprised of the 125 accredited U.S. medical schools; the 16 accredited Canadian medical schools; more than 400 major teaching hospitals; 86 academic and professional societies representing 87,000 faculty members; and the nation's 67,000 medical students and 102,000 medical residents.

### **Services Provided by the AAMC**

The Association of American Medical Colleges (AAMC) provides a number of services to its member medical schools and to medical school applicants and

students. The goal of these services is to promote quality in all areas of academic medicine. These services include:

The Liaison Committee on Medical Education (LCME). The LCME is the accrediting body for U.S. medical education programs leading to the M.D. degree. The American Medical Association (AMA), the professional association for practicing physicians, is a co-sponsor of the LCME.

MEDLOANS. MEDLOANS is a comprehensive loan program that provides financial assistance for enrolled medical students.

The National Residency Matching Program (NRMP). The National Residency Matching Program is a computerized program managed by the AAMC that matches candidates for training in medical residency programs at teaching hospitals throughout the United States according to the candidates' preferences and the needs of the residency programs. For 1996, there were 20,500 residency positions available in 3,400 specialty programs at participating hospitals. 24,700 medical school graduates participated in the Match. Of these, 8,100 were graduates of foreign (non-U.S. or non-Canadian) medical schools.

The American Medical College Application Service (AMCAS). AMCAS is a centralized medical school application service that enables applicants to file a single, standardized application with the AAMC. The AAMC reviews and verifies information on the application, including undergraduate college transcripts, course grades, and admission test scores, and forwards this standard data to any of the 110 participating U.S. medical schools to which the applicant wishes to apply. For the 1995 medical school entering class, 46,591 applicants applied for the 16,253 first-year positions in the 125 U.S. medical schools. Most of these applications were processed through AMCAS, and, on average, each applicant applied to 12 medical schools.

The Medical College Admission Test (MCAT). The MCAT is a standardized test used to assess medical school applicants' science knowledge, reasoning ability, and communication and writing skills. The MCAT was developed by the AAMC. The test is administered twice each year at 600 testing sites throughout the U.S. and Canada and at ten locations outside North America. During the two administrations in 1995, more than 67,000 persons sat for the examination. MCAT scores are one of a number of factors considered by medical school admissions committees at each school in evaluating and selecting applicants.

## **The Admissions Process at U.S. Medical Schools**

Each medical school operates independently when evaluating applicant files and making admissions decisions. While most schools require applicants to apply

for admission by using the AAMC 's centralized application service (AMCAS), many schools also require that applicants complete and submit school-unique, supplemental application forms which provide additional personal data and information. The AAMC does not make admission recommendations or play a role in the admission decisions of the schools. Most medical schools have established admissions committees made up of faculty and staff members. These committees review applicants' files and evaluate applicants' qualifications based upon criteria developed by each school. Potential candidates for admission may be invited to come to the school for an interview.

This entire process may last a year. For example, a college student who plans to graduate from his or her undergraduate college studies in June of 1997 and who wishes to apply for a medical school's entering class, which begins in September of 1997, most likely took the Medical College Admission Test (MCAT) in April 1996. Scores for this test were reported to the student in June 1996. If the student was dissatisfied with his or her scores, the student may have taken the MCAT again at its next administration in August 1996. The AMCAS and the medical schools which do not participate in AMCAS began accepting applications for the September 1997 entering class in June 1996. While the cut off date for receipt of applications vary by school, most schools have application receipt deadlines between October 15 and November 15. During the period from the receipt of applications until March 1997, admission committees at the schools evaluate applications and applicants, conduct interviews, and make final decisions. By March 1997, all applicants will have received an offer of enrollment from or been rejected by each school they have applied to. Some schools place qualified students, who have been rejected due to limits on enrollments, on a waiting list. Applicants on a waiting list may be offered admission in place of accepted applicants who choose to attend another school.

Admission testing has generated substantial public debate and many people have expressed concern about the fairness of standardized test, about test bias, and about over reliance on multiple-choice test scores in admission decision making. As will be indicated later in this article, data has shown that MCAT scores provide useful information about medical school applicants' subsequent academic performance. A 1993 survey of medical school admissions officials revealed that MCAT data is used to:

- Identify applicants likely to succeed in medical school and those likely to experience academic difficulty,
- Assess applicants' strengths and weaknesses in knowledge of entry-level science content, science problem solving, critical and analytical thinking, and written communication skills,

- Interpret grade transcripts and letters of evaluation from unfamiliar undergraduate institutions.

Educational test and measurements experts tell those in the school admissions field that standardized test scores should never be the sole basis for student selection. The AAMC agrees with this philosophy and advises medical school admissions officials that MCAT scores are intended to be only one of several measures of an applicant's academic qualifications that they should consider. Many medical school admissions committees evaluate MCAT scores in conjunction with these other sources of information about the applicant:

- Undergraduate and postgraduate courses' grade point averages (GPA),
- Breadth and difficulty of undergraduate course work,
- Quality of the degree-granting undergraduate institution,
- Letters of evaluation and recommendation from undergraduate advisors, faculty members, and others,
- Involvement in extracurricular activities, such as student government and community service,
- Involvement in and quality of health related work and research experience,
- Participation in other activities demonstrating motivation, responsibility, maturity, integrity, resourcefulness, tolerance, perseverance, dedication to service, and other relevant characteristics,
- Medical school interview results, and
- State or county of legal residence in the United States.

### **The History and Development of the Medical College Admission Test (MCAT)**

The AAMC first sponsored an objective test for applicants to medical school in 1930 for the purpose of reducing the high attrition rate among entering freshman. This test was called the Scholastic Aptitude Test for Medical School, and versions of it were used until 1946. At that time a new test, the Professional Aptitude Test, was introduced. In 1948, the name of this test was changed to the Medical School Admission Test (the MCAT). Since that time, the MCAT's content, format, and score reporting conventions have evolved and changed to meet the needs of medical schools to select and prepare students for the chang-

ing requirements of medical practice as the knowledge base and technologies of medicine rapidly change and expand.

The current version of the MCAT was introduced in 1991. It is the result of revision efforts started in 1983 with the formation of the Ad Hoc Advisory Committee to the MCAT Essay Pilot Project which was charged with determining the utility and feasibility of including a writing assessment as part of the MCAT. In addition, revision of the MCAT's multiple-choice test sections began in 1987, guided by the MCAT Evaluation Panel. Both committees were composed of AAMC constituents and staff. The two committees concluded their work by submitting a single set of recommendations which set forth their conceptual framework which would underlie the format of the 1991 revised test battery. These recommendations reinforced the medical community's view that physicians must be more than holders of vast amounts of scientific information. They emphasized the need for future physicians to have the ability to gather and assess data, to apply the basic concepts and principles of medicine to the solution of scientific and clinical problems, to evaluate situations critically, and to arrive at logical solutions. Physicians also must update their knowledge and skills continually and be able to communicate effectively with patients, colleagues, and the public. Mastery of basic concepts in biology, chemistry, and physics, while still considered prerequisite, was not judged to be a sufficient indicator of success in medical school.

### **Composition of the Current Version of the MCAT**

The revised MCAT was designed to reinforce medical education's call for candidates with balanced undergraduate preparation and with knowledge of basic science concepts as well as the ability to think critically. Its three multiple-choice sections (Verbal Reasoning, Physical Sciences, and Biological Sciences) and directed writing assessment (the Writing Sample) were developed to assess (1) mastery of basic concepts in biology, chemistry, and physics, (2) facility with scientific problem solving and critical thinking, and (3) communication and writing skills. To discourage premedical students from concentrating too heavily in science course during their undergraduate education, science content coverage on the MCAT was restricted to the topics and subject matter taught in first-year biology, chemistry, and physics courses at most colleges.

The MCAT is composed of four distinct sections. The test is 5 3/4 hours in duration. Two sections are administered in the morning followed by the remaining two sections after a lunch break. Each section is scored separately and a distinct, scaled score is reported for each section. The sequence of each section and allocated times are as follows:

| Section             | Number of Questions | Time in Minutes |
|---------------------|---------------------|-----------------|
| Verbal Reasoning    | 65                  | 85              |
| (break)             |                     | (10)            |
| Physical Sciences   | 77                  | 100             |
| (lunch)             |                     | (60)            |
| Writing Sample      | 2 topics            | 60              |
| (break)             |                     | (10)            |
| Biological Sciences | 77                  | 100             |

### **Verbal Reasoning**

The Verbal Reasoning section of the MCAT is designed to assess the examinee's ability to understand, evaluate, and apply information and arguments presented in prose texts. The test consists of several passages, each 500 - 600 words long, taken from the humanities and social sciences and from areas of the natural sciences not tested on the MCAT Physical and Biological Sciences sections. Each passage is accompanied by six to ten multiple-choice questions based on the information presented in the passage. Some questions assess basic comprehension of the text; others require the examinee to analyze data, evaluate the validity of an argument, or apply knowledge gained from the passage to other contexts. Since the humanities, social sciences, and natural sciences include a vast range of subjects and since courses in these areas differ greatly in content, test questions do not cover a specific set of topics. Examinees are not tested for specific subject knowledge in the disciplines addressed by the Verbal Reasoning section.

The questions on the Verbal Reasoning test measure skills in the four cognitive areas of: comprehension, evaluation, application, and incorporation of new information.

### **Physical Sciences**

The MCAT Physical Sciences section is composed of multiple-choice questions that assess an examinee's reasoning skills in general chemistry and physics. The majority of questions are based on passages, each about 250 words in length, that describe a situation or a problem. The test consists of ten or eleven of these problem sets, each followed by four to eight questions. An additional 15 questions are independent of any passages and of each other. Neither the

passage related questions nor the independent questions test the ability to memorize scientific facts. Rather, both types of questions assess the examinee's knowledge of basic physical science concepts and facility at problem solving in physics and physically-related chemistry areas. Many passages also include data presented in graphs, tables, and figures, and some of the questions related to these passages require interpretation of the data. The incorporation of items requiring examinees to understand and interpret data presented in graphs, charts, and tables with text material that crosses science disciplines provide a rich basis for assessing problem solving on the MCAT.

The concepts covered by the Physical Sciences section were determined by surveying undergraduate faculty members on the level of coverage given to potential topics in the introductory chemistry and physics course sequences. Medical educators, medical students, and physicians were asked the relevance and importance of potential topics for the study and practice of medicine. Chemistry and physics topics judged as prerequisite by medical school respondents and covered at a majority of undergraduate institutions in the introductory science course sequences were selected for inclusion in the test. The chemistry and physics concepts included in the test are considered basic. Advanced level undergraduate course work in physics and chemistry is not required to successfully take the test. A knowledge of calculus is not required. The mathematics concepts required for the Physical Sciences section are those typically required of students in introductory science courses. Topics included in the Physical Sciences test section include:

|   |                                     |  |
|---|-------------------------------------|--|
| Acids/Bases                                 | Atomic and Nuclear Structure        | Bonding                                  |
| Electrochemistry                            | Electrostatics and Electromagnetism | Electronic Circuits                      |
| Solution Chemistry                          | Equilibrium and Momentum            | Light and Geometrical Optics             |
| Fluids and Solids                           | Force and Motion, Gravitation       | Phases and Phase Equilibria              |
| Stoichiometry                               | Translational Motion                | Sound                                    |
| Electronic Structure and the Periodic Table | Work and Energy                     | Wave Characteristics and Periodic Motion |
| Rate Processes in Chemical Reactions        | Kinetics and Equilibrium            | Thermodynamics and Thermochemistry       |

### **Writing Sample**

By requiring candidates to develop and present ideas in a cohesive manner, the MCAT Writing Sample offers medical school admission committees evidence of applicants' writing and analytical skills. These skills are critical to the preparation of useful medical records and to effective communications with patients and other health professionals. The MCAT Writing Sample provides unique information unavailable from other sections of the examination.

The Writing Sample consists of two essays, each of which the examinee is allotted 30 minutes to

write. The Writing Sample is designed to assess skills in the following areas:

- Developing a central idea
- Synthesizing concepts and ideas
- Presenting ideas cohesively and logically
- Writing clearly following accepted practices of grammar, syntax, and punctuation consistent with timed, first-draft compositions.

Each essay question provides a specific topic that requires an expository response. Essay topics do not pertain to the technical contents of biology, chemistry, physics, or mathematics; the medical school application process or the reasons for the choice of a medical career; social or cultural issues not in the general experience of MCAT examinees; or religious or other emotionally charged issues. Each of an examinee's essays is rated on a six-point scale by two separate readers. If scores for an essay vary by more than a point, a third reader re-evaluates the essay. The total scores for both essays are summed and converted to an alphabetical scale for reporting. Medical schools receive the alphabetical score and, at their option, copies of their applicants' essays.

### **Biological Sciences**

The format for the MCAT Biological Sciences section is similar to that of the Physical Sciences section. Multiple-choice questions are used to assess reasoning skills in biology and organic chemistry. The test includes ten or eleven problem sets with four to eight questions each and 15 questions independent of any passage and of each other. Passages are approximately 250 words in length. Some questions require interpretation of information found in graphs, tables, or figures. Questions do not assess rote memorization of scientific facts; instead they test knowledge of basic biological sciences concepts and problem-solving abilities in biology and the biologically-related areas of chemistry.

Like the concepts in the Physical Sciences section, biological sciences concepts were identified by surveying undergraduate and medical school faculty, medical



students, and physicians. Biology and organic chemistry concepts included in the Biological Sciences section are considered basic. They are taught at the introductory level at the vast majority of undergraduate institutions. Advanced undergraduate-level course work in biology or organic chemistry is not required for the test. The following topics are covered by this section:

|  |  |  |
|--|--|--|
| Amines   | Biological Molecules                             | Digestive and Excretory Systems          |
| Evolution  | Hydrocarbons                                     | Generalized Eukaryotic Cell              |
| Genetics   | Skin Systems                                     | Muscle and Skeletal Systems              |
| Microbiology                                       | Oxygen-Containing Compounds                      | Nervous and Endocrine Systems            |
| Molecular Biology: Enzymes and Cellular Metabolism | Molecular Biology: DNA and Protein Synthesis     | Organic Covalent Bonding                 |
| Respiration Systems                                | Separations and Purifications                    | Reproductive Systems and Development     |
| Circulatory, Lymphatic, and Immune Systems         | Use of Spectroscopy in Structural Identification | Specialized Eukaryotic Cells and Tissues |

The MCAT provides admissions committees with a standardized measure of academic achievement for all examinees. Differences in undergraduate curricular emphases, evaluation standards, and grading scales make some preadmission information, such as undergraduate grades from different colleges, difficult to interpret. The MCAT provides assessment information about specific knowledge and skills on a standard scale for all examinees.

### Test Construction and Discrimination Reviews

The test questions used in the multiple-choice sections of the MCAT (Verbal Reasoning, Physical Sciences, and Biological Sciences) are designed and written individually by college and university professors from throughout the United States. These are experienced teachers in the fields of chemistry, physics, biology, and the humanities who have contracted with AAMC to confidentially develop and write MCAT test items. Each item writer is assigned a specific subject area and skill level which his or her test passages and questions must examine. Submitted questions are initially reviewed by our in-house editors. The test questions are then forwarded to external subject matter experts for

technical review and to prominent and educators who examine the questions for potential gender and racial or ethnic bias or discrimination.

After test questions have been thoroughly reviewed, edited, and approved for inclusion in the MCAT, each passage and its related questions are pre-tested as field test questions. In each operational MCAT test form which is administered, at least one passage and its related questions is a field test question set. These questions are not scored. Test takers know that such questions are included in their test, but they do not know the position or location of the field test questions. Subsequent to a test administration, each test question undergoes intense statistical analyses. Degree of difficulty for each question is calculated based on student performance on the question. Differential results by students' gender and racial/ethnic backgrounds are examined to identify any previously undetected question bias

Following all these reviews and analyses, passages and questions are approved for inclusion in an operational MCAT examination as scorable test items. Questions are assembled into operational test forms based on the subject matter and skill levels each question assesses and on each questions degree of difficulty as determined during field testing. This process ensures that each test form assesses the knowledges and skills required by the MCAT Test Specifications and that different versions of the MCAT have comparable levels of difficulty.

### **Predictive Validity Research Study for the MCAT**

Concurrent with the introduction of the revised MCAT in 1991, a plan for assessing the predictive value of the new MCAT was designed. Fourteen medical schools were selected to participated in the study, and they agreed to provided the necessary data. These schools were chosen to represent the 125 U.S. medical schools. The 14 schools are located in different regions of the country; some are privately operated and the remainder are state supported; the selected schools have a proportional mixture of different types of curricula; and the size and racial or ethnic composition of their student bodies are representative of most of the other U.S. medical schools.

The study is following two cohorts of students comprised of those who entered the selected medical schools in the fall of 1992 and in the fall of 1993. In the U.S., the academic school year at most schools begins in August or September and ends in May or June. Students in these two cohorts were identified, and preadmission data (data which was available to the medical schools' admissions committees when their admission decisions were made) were gathered on the students. Preadmission data obtained for each student consisted of undergraduate grade-point average (UGPA), MCAT scores, and the selectivity index for

their undergraduate institution. The undergraduate grade point average is a numerical grade used in the U.S. which is the average of all of a student's college course grades weighted by course credit hours. The selectivity index is a quality index number assigned to all undergraduate colleges and universities in the United States based upon the mean college admission test scores of their students enrolled in a given year. These data allowed evaluation of the MCAT's predictive value within the context of other data typically used during admissions decision making.

To date, performance data for the 1992 entering class at the selected medical schools has been gathered. The students' performances were summarized by computing end-of-year grade-point averages (GPAs) for each year (year 1 of medical school and year 2 of medical school) and a two-year cumulative grade-point average which was the simple average of the two end-of-year averages.

The extent to which MCAT performance predicts performance on Step 1 of the United States Medical Licensing Examination (USMLE) was also examined. The licensing exam is administered in "steps" during the course of a medical student's education and training. All three Steps of the USMLE must be passed before a medical school graduate is licensed to practice medicine in the United States. Step 1 of the USMLE is taken by most medical school students during their second year of medical school.

Before analyzing the relationships between preadmission variables and medical school grades, descriptive information for each of the preadmission variables were obtained. The data were not collapsed across the schools; instead, the distribution of each of the preadmission variables was investigated within each of the schools.

The relationships among the MCAT, Undergraduate GPA, institutional selectivity, medical school grades, and Step 1 scores were investigated using multiple regression analysis. Regressions were run separately for the entrants to each school. Because applicants with low values on preadmission variables are typically not selected for medical schools, the range of values for preadmission variables is generally restricted or truncated. Corrected correlations estimate the strength of the relationships between the preadmission data and the medical school performance variables in the absence of selection. The restriction-in-range corrections were based on MCAT, UGPA, and selectivity data for each school's applicant pool.

Multiple correlation coefficients were obtained for five unique predictor sets, and the results were summarized across schools. The five predictor sets were:

Undergraduate GPAs

MCAT scores

Undergraduate GPAs + selectivity index

Undergraduate GPAs + MCAT scores

Undergraduate GPAs + selectivity + MCAT scores

The following table reports the ranges and median values for the corrected multiple correlation coefficients obtained for the five sets of predictors. As is evident in the table, MCAT scores appear to have a slightly higher correlation with medical school grades (median correlations ranging from .62 to .67) than do UGPA data (median correlations ranging from .54 to .58). However, prediction of performance was improved when the two sets of predictors were considered jointly (median correlations of .70 to .76).

Range of Median Values of the Corrected Correlations between Preadmission Data Sets and Medical School Performance for the 1992 Entering Class at 14 Schools

| <u>Predictor Sets</u>                           | <u>Medical School Performance</u> |                   |  |                               |
|---|-----------------------------------|-------------------|--|-------------------------------|
|   | <u>Year 1 GPA</u>                 | <u>Year 2 GPA</u> | <u>Cumulative<br/>Year 1&amp;2 GPA</u> | <u>USMLE<br/>Step 1 Score</u> |
| <b>Undergraduate GPAs</b>                       |                                   |                   |  |                               |
| Median  | .54                               | .58               | .58                                    | .48                           |
| Range   | (.40 - .74)                       | (.23 - .74)       | (.42 - .73)                            | (.35 - .64)                   |
| <b>MCATs</b>                                    |                                   |                   |  |                               |
| Median  | .67                               | .62               | .64                                    | .72                           |
| Range   | (.38 - .78)                       | (.46 - .72)       | (.44 - .78)                            | (.58 - .79)                   |
| <b>Undergraduate GPAs + Selectivity</b>         |                                   |                   |  |                               |
| Median  | .60                               | .61               | .64                                    | .53                           |
| Range   | (.50 - .80)                       | (.29 - .79)       | (.49 - .82)                            | (.40 - .73)                   |
| <b>Undergraduate GPAs + MCATs</b>               |                                   |                   |  |                               |
| Median  | .75                               | .70               | .76                                    | .75                           |
| Range   | (.48 - .80)                       | (.61 - .77)       | (.63 - .81)                            | (.62 - .81)                   |
| <b>Undergraduate GPAs + MCATs + Selectivity</b> |                                   |                   |  |                               |
| Median  | .75                               | .73               | .76                                    | .75                           |
| Range   | (.53 - .82)                       | (.64 - .81)       | (.63 - .84)                            | (.62 - .82)                   |

A substantial difference is evident when comparing median correlations for MCAT and UGPA data with USMLE Step 1 performance. Here, MCAT is much more strongly related to Step 1 (median value of .72) than UGPA (median value of .48). Again, however, when MCAT and UGPA data are considered jointly, the median correlation coefficient increases.

The MCAT's utility in the admission process becomes evident when values obtained for predictor set 3 (UGPA + selectivity) are compared with those ob-

tained for predictor set 5 (UGPA, selectivity, and MCAT). The median values of the prediction correlations increase notably when the MCAT scores are added to the equation.

Because the MCAT's primary purpose is to identify individuals most likely to succeed in medical school, initial efforts have focused on the evaluation of the predictive validity of the test. This study to date represents a comprehensive evaluation of the predictive validity for the first two years of medical school, which in the United States are heavily weighted towards the study of the basic sciences. Establishing the predictive validity of the MCAT for these first two years of basic medical science study is a crucial step in the overall validity studies planned for the MCAT. Data are being collected from the entering class of 1993 to replicate the present study's results for the 1992 entering class. Subsequent studies will examine correlations with performance in the clinical years of medical school training as well as performance on the other, sequential Step Examinations that comprise the U.S. Medical Licensing Examination process.

## **Summary**

The decision to select and to admit an applicant to a medical school in the United States is one that is made individually by each medical school according to each school's own criteria. Multiple factors are reviewed and considered by medical school admission committees when making admission decisions. One of these factors is the applicant's performance on the Medical College Admission Test (MCAT). The current version of the MCAT was specifically designed to assess the skills identified by medical school faculty and physicians as those necessary for a student to succeed in medical school and in the practice of medicine. The results of MCAT predictive validity studies conducted to date indicate that the MCAT is a strong predictor of performance in the critical first two years of medical school. Additional studies are planned which will extend the examination of the MCAT's correlation with performance into the third and fourth years of medical school students' clinical studies and to their performance on the final stages of the licensing exam process.

## References

Material in this paper was adapted or extracted from the following sources:

Leadership for Academic Medicine. Washington, DC: Association of American Medical Colleges, 1996.

MCAT Student Manual. Washington, DC: Association of American Medical Colleges, 1995.

Use of MCAT Data in Admissions: A Guide for Medical School Admission Officers and Faculty. Washington, DC: Association of American Medical Colleges, 1994.

Mitchell, K., Haynes, R., and Koenig, J. Assessing the Validity of the Updated Medical College Admission Test. *Academic Medicine* 69(1994): 394-401.

Wiley, A., and Koenig, J. The Validity of the Medical College Admission Test for Predicting Performance in the First Two Years of Medical School. *Academic Medicine* 71 (Oct Suppl. 1996): S83-S85.

## **Admission to the study in medicine in Belgium: two 'different' solutions to the 'same' problem; reflections of a Flemish school psychologist**

**P. J. Janssen**

A few weeks ago a specialized Flemish magazine for general practitioners (the so called 'Artsenkrant') announced that Belgium at this very moment disposes of 35.000 family doctors and medical specialists. That means an average of one medical doctor for every 280 inhabitants. Annually about 1200 students complete their seven years of medical basic training; from the year 2004 only 700 of them will be allowed - as was decided by the federal ministry of National Health about a year ago - to enter the profession. The problems in the domain of dentistry are analogous and need to be solved in the same time. That's the problem my country has to solve to the best of its possibilities, being obliged to do so for several reasons: to control the costs of the national health insurance, to maintain the quality of medicine and dentistry on an accepted level (as the organizations of practitioners were claiming). Due to the way the Belgian State internally is organized - now going its way into a federation - this 'central' decision has to be elaborated within each of its both 'national' governments, responsible for their educational systems as such. That explains way this 'same' problem - within 'one' but subjectively not 'the same' country - could create two 'different' solutions.

In order to enable full understanding of this problem from the Belgian perspective, I start my contribution by presenting some additional information about my country (1). Next I will present the two solutions for the problem as they will be elaborated from now on (2). Being personally involved in the development of the selection procedure, as it will be applied in the Dutch speaking part of my country, I will present you the arguments of the school psychologist I am, for preferring this Flemish approach (3).

### **1. The Belgian context**

As many of you know, Belgium is from its origin in 1831 a parliamentary democracy with a constitutional hereditary monarchy. The governmental system set up by the constitution of 1831, which vested most powers in the central government, was modified by the division of the country into linguistic areas in the 1960s. Constitutional revisions in the 1970's and 1980's resulted in the establishment of a quasi-federal governmental structure based on the three official linguistic communities: Dutch (about 57 % of population), French (42 %) and German (1 %). These got their own 'national' governments and councils, which acquired a considerable degree of influence over regional and national affairs.



One of their domains of autonomy refers to education. One of the implications was already mentioned: the 'national' problem of too many medical doctors will be solved by at least two different approaches.

To understand the way the problem is posed, one also has to know that nearly all Belgians are covered by national health insurance, which as such is, at least until now, managed for Belgium as a whole, thus on the federal level. This social security system also covers work accidents, unemployment, premature death, occupational diseases, invalidism, and retirement pensions. The financial cost of such a system is enormous. The number of medical doctors is one of the system's parameters, which have to remain controllable. Now one understands why the central level could formulate a 'problem' that needs to be solved on both national sub-levels...

One of many other common Belgian problems refer to the transition youngsters make from secondary to tertiary education. The latter implies university as well as the so called Higher Education outside university. Both subsystems are freely accessible for every youngster who has completed her and his general secondary education. There are only two exceptions: the one refers to the five year university training in engineering (within faculties of applied sciences); the other to some fields within the artistic Higher Education outside universities. For both a freshman has to pass an entrance exam assessing competency, resp. in mathematics or artistic potential. Everyone who passes, has the right to register as a student. Until now there is no numerus clausus or fixus. Important to know also is that Belgian universities are subsidized by their government on the base of enrollment figures: the more students (and the more freshmen), the 'better', at least at the beginning of the academic year... Also important to know is (1) that Belgium as such has no common 'national' achievement test or exam (some type of 'Abitur') at the end of secondary education; institutions themselves (also subsidized on the base of enrollment numbers !) autonomously award the degree and ...(2) that there exist real differences between these institutions, also as far as educational levels attained. The consequence of this quite 'complex' and at the same time quite 'generous' system are, at least, twofold:

(1)The first refers to the quite low success rates at the end of the first year, where the student - as he has to do every year during the curriculum - has to pass a fixed number of exams, each referring to a course he is obliged to take. Decisions are taken by a jury, grouping the professors who taught these courses. The student who is unsuccessful at the end of June, may try again at the end of August. The success rates at the end of the first year - both exam periods combined - vary from 1 out of 3, to 1 out of 2. Relatively most successful are students in engineering (about 65 %), immediately followed by students in medicine, what proves, as a matter of fact, that on the average this faculty already attracts capable and well motivated students.

(2) The second consequence has to do with the attractiveness of this system of free entrance for young Europeans, who, being not allowed to the study of their choice by restricted entrance regulations within their own country, detect Belgium (as 'the land of hope and glory' ?) where they can start the study they want. So Flanders already for many years 'had to accept' and 'to pay' for many hundreds of students from the Netherlands, who were eliminated from medicine by the system of lottery in their own country. Some years the Dutch freshmen even outnumbered their Flemish colleagues in the Antwerp university ...

The basic curriculum in medicine counts 3 years of 'candidature' plus 4 years of second cycle. The study program of the very first year enables freshmen to master the necessary basis in mathematics, physics, chemistry and biology. During the second and third year anatomy and physiology are extensively studied. During the second cycle students get their specific medical basic training. Belgian medical faculties have their own educational climate. At the end of these 7 years of study students are 'selected' for specialization on the base of study results obtained during the preceding six years. This rule may explain why students in medicine prove to be hard working, and most of the time quite competitive, which in turn may explain some 'negative' behaviours of some students with respect to materials - for instance in libraries - which also might be of interest to their fellow-students. This 'third cycle' of specialization varies in time; it can take up to 5 years.

In 2004 the federal government will allow only 700 new medical doctors to enter the national health system. That year was chosen in order to offer both national communities sufficient time to take the measures they judge essential. In the case this measure had to be applied immediately, about 500 'doctors' out of the 1200 as 'delivered' this year, had to find another job ...

## **2. Different solutions to the 'same' problem**

That is the problem my country was confronted with about a year ago. The two national Ministers of Education - Grafé in the French-speaking and Van den Bossche in the Flemish community - had to develop an option. In the case one would apply a division according to the percentages within the population, this would imply a reduction to about 399 doctors in Flanders and 301 in Wallonia . Different measures can be taken: one can introduce them at the end of the curriculum, apply them at the moment of transition from the first cycle into the second (after three years of study), take them at the end of the first year, or do it before students can register. Within each community discussions started immediately.

## **2.1. The 'French' way of solving (or postponing ?) the problem ...**

In the French-speaking part the decision was taken to postpone this selection to the end of the first study cycle. So at least everyone, knowing what will happen, may try his chance. After three years only a limited number of students will be allowed to continue in medicine. The others - if there are more students than needed - have to find another study domain. For the moment it is not clear - at least not for me - how these future medics will be selected. It can happen on the base of their study results; in that case, I guess, one needs some type of national test, because these students study in different universities (Liège, Bruxelles, Louvain-la-Neuve). Another possibility would include that they, as 'candidates in medicine', have to apply for some type of a lottery system.

## **2.2. The solution in Flanders**

In the Dutch-speaking part of Belgium the discussion about measures to be taken quite soon became dominated by the decision taken by the Minister of Education, Luc Van den Bossche, member of the Flemish Socialist Party. He decided to introduce a 'national' entrance selection before the beginning of the academic year 1996/1997. Every youngster wanting to study medicine has to take part; only those who will succeed, will be allowed to choose a university and to register as a freshman. A working group was charged on December 1, 1995, with the task to design and implement a procedure, which should be constructed in such a way, that future students will not be selected on the base of their educational attainment in Secondary Education, but 'only' on the base of their real medical aptitude, including also their motivation. ...

### ***2.2.1. The procedure as developed in the beginning of 1996***

This working committee consisted of 21 members, experts (from the Ministry of Education, university professors lecturing in medicine and psychometrics) as well as representatives of students in medicine. These people met about 10 times under the dynamic direction of Jan Adé, the Director of Higher Education and Scientific Research in the Flemish Ministry of Education. On January 26 the commission could meet Dr. Günter Trost, who described the German TMS and the way selection for medicine is functioning in Germany.

In the beginning of May 1996 - at the end of this same month the Flemish Parliament had to decide about a pro-position concerning the introduction of this entrance test - the Commission presented its report to Minister Van den Bossche. This document describes the work done and motivates the options chosen after a thorough evaluation of the pro's and con's of all kinds of measures. Interviewing of all candidates - about 1000 youngsters are expected to take part in the entrance examination - turned out to be for more than one reason completely impossible. The same holds for the so called 'apprenticeship in

nursery', enabling a short period of work in a hospital, which the representatives of the students judge to be an essential component of the procedure in order to enable a - at least subjectively - valid test of a youngster's motivation for the medical profession. Even a test for manual dexterity had to be excluded for students in dentistry, who also had to participate in this entrance examination.

The resulting test will consist of two big parts, each taking a full 'day' of testing. (1) In the first 'Knowledge of and Insight in Sciences' (KIS) will be tested by means of four 'different' instruments, resp. assessing an entrants competency in the domains of mathematics, chemistry, physics and biology. In each of these the content will be defined in terms of the minimum program Secondary Education is offering; the items will be constructed in such a way that they 'test' productivity - i.e. the constructive use of knowledge and insight - rather than pure reproduction. These tests proved to be essential, because Belgium does not, as other European countries do, test achievement at the end of Secondary Education by means of an uniform and common final exam. (2) During the second day an applicant's capability in the Acquisition and Elaboration of Information (the AEI-part of the exam) will be tested. Here instruments have to be developed to measure inter-individual differences concerning cognitive functioning, memory, assimilation of visual information and last but not least the ways these candidates handle a casus, describing, by means of modern audio-visual media, the complexity of a real-life medical problem from the perspective of the patient as well as the 'different' medical experts involved on the base of the information each of these disposes of. The idea for this last test was developed by my colleague Prof. Paul Coetsier from the university of Gent on the base of an expertise already available in the domain of selection of managers for industry.

Given the fact that the Flemish Minister of Education initially obliged the commission to organize this entrance exam - as a first try-out - twice before the start of the academic year 1996/1997, the commission simultaneously had to develop a strategy concerning the decisions to be taken on the base of the results of this two-days testing program. As a consequence two decision rules were formulated:

- a) The first reflected the criteria for grading as they are used in Flemish universities. Everyone who reaches the criteria put forward, will receive at least a score of 10 out of 20 points. This score will be as higher as the level on which this performance can be placed, exceeds these minimal objectives; as such a student can receive distinction, great distinction or the greatest distinction, in the case his level of attainment reaches a core of 14/20, 16/20 or 18/20. Given the fact that not all parts of KIS are equally well taught in Flemish Secondary Education, the commission decided to take only each applicant's three 'best' scores into consideration. As far as AEI is concerned the same criterion holds.

By consequence, in order to succeed in this entrance exam, an applicant has to reach in each of the two parts of this entrance test at least a score of 10/20.

- b) The commission decided also to allow - given this very first application of the procedure - also those youngsters, who, after the combination of their main scores on the two parts of this exam, obtained at least a T-score of 40 win the total group of applicants. This decision would imply that at this very first try-out of the entrance procedure 'only' about 15 % of the applicants would be not successful at all. So quite a safe start could be made, given the political considerations that urged Minister Van den Bossche to take this measure.

In the beginning of June the Flemish Parliament accepted this proposal. At the same moment, the execution of the decision taken was postponed until the start of the academic year 1997/1998 ... The reason was quite acceptable; it turned out to be too late to inform last year students in Secondary Education about the implications this decision would have on their vocational planning ...

### ***2.2.2. Procedure as to be applied in 1997***

In the same time it became clear that the first application of the parliamentary decision will take place at the beginning of the next academic year. In the way this Flemish parliamentary decision was taken, it implies some and even important changes from what the working group proposed to the Minister with respect of a very first application before the start of the academic year 1996/1997. The idea of a two days exam will hold. There will be, according to the Belgian academic tradition, two exam periods, implying so a second chance for those who did not succeed at the first occasion. Not well defined are both points in time involved: From a traditional university point of view the beginning of July and September seem appropriate; taking the perspective of a school psychologist, by definition well informed about the ways last year students in Secondary Education are going through their process of vocational decision making, two other moments seem to be more appropriate: the period after Eastern and the end of the school year in the beginning of July. The minister has to take his decision in one of the forthcoming weeks.

Between times it became clear that in the decision, as it was taken by the Flemish Parliament, the norm to succeed has been changed drastically as compared to the one proposed by the initially installed working group. In order to succeed, an applicant has to obtain now for each of the two constituent parts a score of at least 12/20. In the forthcoming days the jury - under the direction of chairman Jan Adé of the commission that prepared the procedure - officially will be installed. From that moment on, the way the entrance exam, as it will take place within the Flemish Community, will get its definite content and form. It will be organized by an already existing governmental service - the

‘Permanent Recruitment Secretariat’ (‘Vast Wervingssecretariaat’) - which disposes of the necessary material and logistic infrastructure.

### **3. A school psychologist’s evaluation**

For the moment I see one essential aspect of the problem, which as such has not yet appropriately been analyzed. Until now the Flemish Minister Van den Bossche refuses to define the contingent of students to be allowed to start next year their studies in medicine. This means an extra complication for the commission who will be in charge of the preparation of the entrance examination and for the jury which has to decide about each individual applicant.

*The idea of an entrance examination as such is in my view a sound approach towards the solution of a problem which - at least as I see it - is broader than this specific case in medicine and dentistry.* As already mentioned, Belgium is confronted with quite low, even too low, success rates at the end of the first year in Higher Education. At least as far as Flanders is involved, I myself already proposed in 1988 (Janssen & De Neve, p. 179 - 183) a set of measures, which some years later was taken over by a Flemish group of experts in university education ( the so called ‘Contactgroep Academic Onderwijs’) (CgAO 1993). These include the improvement of the ways last year students in Secondary Education have to be counseled in the making of their vocational decision and the didactic of the first year in university. Included within this counselling process are two series of tests, to be taken individually at two different moments during the process of vocational decision making. The second one could be replaced by an entrance selection, as it is proposed here for the studies in medicine and dentistry.

The proposal itself is based on the two big causes for these transition problems, as both are identified now on the base of about half a century of scientific research. The one refers to the lack of competence of entrants in Higher Education (3.1); the other to the quite defective self regulation of freshmen in Higher Education (3.2). Although both problems turn out to be, to a certain degree, independent (even ‘competent’ students do not succeed), they can be solved together by means of, as was already said, the improvement of student guidance during the last year in Secondary Education (3.3).

#### **3.1. Competence: necessary but as such insufficient**

An analysis of the process of studying makes clear that a serious lack of competence creates a break-down in the progression of studying as a process. As such studying has to be defined (Janssen, 1996) as the integration of learning and thinking. Both components can be described separately; so it becomes quite clear what their integration has to imply.

**Learning** has to be meaningful and comprises, according to Ausubel (1963), within a positive feedback loop the three stages of assimilation, transformation and accommodation. Assimilation implies complete understanding - on the base of available foreknowledge - of what is presented in lectures. Transformation implies intelligence, because the student himself has to change his until then appropriate cognitive structure into a new logically coherent 'Gestalt', which will allow him, having put this new one into memory, accommodation, i.e. the expertise to solve the problems he will be confronted with as the expert he wanted to become.

The impact of **thinking** was already mentioned in my description of the transformation the student, as a learner, has to achieve. But it also works, as can be seen within a moment, at both other stages involved. Thinking can be threefold, resp. in line with the so called combination, comparison and cause-effect schemes. The first implies the bringing together of two separate elements into one new whole, the second the understanding of something new on the base of the 'old' one already is familiar with, the third the 'working' of one element in the 'making' of another ... These three - as was demonstrated by Gordon Pask (1976) - cannot work independently from each other, given the complexities students are confronted with.

**Integration** of the three stages of learning with the three ways of thinking creates - as is shown in figure 1 - the going together of up to nine 'different' mental operations, all involved within the process of studying. The loop in learning as well as the interdependencies in thinking remain 'working' within this 3x3 matrix scheme. So up to four loops can be identified within a psychologically well cohering way; each of which - at least in principle - can be tested, also before entering into a specific domain of study in Higher Education: (1) So cognition and memory have to enable the student by means of selective comparing the understanding of the novelties - in mathematics, physics, chemistry, ... - he will meet within a specific study program. (2) By integrating convergent production (being the systematic and quite logic approach of a problem) and 'selective combining' one can test a students capacity in problem solving, reasoning (as such the factoranalytically identified dimension 'reasoning' within the German TMS)... (3) By integrating 'divergent production' (a person's capacity to produce meaningful 'wholes' in making a synthesis) and 'selective encoding' (or that same person's capacity to analyze a complexity by distinguishing the relevant from the irrelevant elements within a given context) one can test a student's capacity for finding meaning within so far new information (it remembers me the dimension 'visual information processing' within the German TMS) ... (4) By integrating organizing and timing, as such quite important for a student's self regulation, one can test in advance a student's capability in a fourth essential component of his studying as a process (as such the psychological meaning of the 'factor' memory as identified within the German

TMS ?) . (+1=5) In the meantime the operation ‘evaluation’ turns out to be the real hinge in this same studying; it refers, as will be described immediately, to the way a student is capable to integrate these different process dimensions into (his experience of) ‘optimal functioning’ in what he is undertaking.

|   |  |  |   |
|---|--|--|---|
| Divergent production<br>(GUILFORD)          | Cognition /<br>Memory<br>(GUILFORD)        | Convergent production<br>(GUILFORD)          | Accommodation<br>(Doing; expertise)                   |
| Organizing                                  | Evaluation<br>(GUILFORD)                   | Timing                                       | Transformation<br>(doing into knowing and vice versa) |
| Selective combining<br>(STERNBERG)          | Selective comparing<br>(STERNBERG)         | Selective encoding<br>(STERNBERG)            | Assimilation of novelty (knowing)                     |
| thinking by means of the combination scheme | thinking by means of the comparison scheme | thinking by means of the cause effect scheme | STUDYING as a process: product of Thinking x Learning |

Figure 1. The nine intellectual operations as involved in the process of studying, implying the integration of the 3 stages as involved in learning (from assimilation into accommodation and backwards) and the 3 ‘schemes’ of thinking, implying the going together - by means of a student’s permanent evaluation - of 4x2 mutually well co-ordinated mental activities.

There is an *alternative way* to test a student’s competence for what he is planning. Instead of confronting him with a series of specific tests, one can offer him the opportunity to study, during at least two hours, a representative sample of study materials from the domain he is interested in. This has the big advantage of offering him a real - at least in his eyes - face-valid test or try-out. He probably will be more inclined to accept the conclusions resulting from that ‘test’ or ‘personal experience’ as to be deduced from it. In line with that principle Minnaert & Janssen (1992) constructed a predictive as well as a nomological valid ‘Excursion into psychology as the domain of my future study’, explaining up to 56 % of variance within study results after a five year follow-up period. Such a type of instrument resembles quite well the casus my colleague Coetsier is constructing for the Flemish entrance test. In the meantime it became quite clear that the student’s study behaviour itself is an as essential component as competency itself turned out to be ...

### **3.2. Self-regulation: necessary but insufficient, when ...**

This study behaviour has to be defined - so the internal logic of figure 2 can immediately be understood thanks to one’s “capacity for finding meaning” (as an aspect of ‘visual information processing’) - as the task-oriented interaction between the student (as a person) and his environment (or the study landscape



he is travelling in), enabling him the achievement of the goal set forward and naïvely attributing its attainment (according to Heider (1958) as the founding father of modern psychological theories of causal attribution) to the product of ‘to can’ (or ‘to be able to’) and ‘to try’ (combining so the ‘what and why ?’ or meaning with the ‘how hard ?’ or working in his doings). So up to nine specific study experiences - as identified in the mean-time by means of factor analysis of item materials of Likert-type study behaviour questionnaires - can find their place within an analogous 3x3 matrix concept. Each of these nine can be psychologically measured in a quite reliable way, offering so the individual student a real life picture of his study method, i.e. his - whether or not - effective regulation of his ‘way-into’ the criteria he has to meet at the end of the year ...

| <b>STUDYING</b><br>as experienced as<br>a students behavior | <b>to try:</b><br>What ? Why ?<br>Meaning | <b>to can:</b><br>to be able to | <b>to try:</b><br>How hard ?<br>Effort |
|---|---|---------------------------------|--|
| <b>Person</b>   | <b>Intention</b>                          | <b>Self confidence</b>          | <b>Activity to control</b>             |
| <b>Task over time</b>                                       | <b>Agenda</b>                             | <b>Effectiveness</b>            | <b>Discipline</b>                      |
| <b>Environment</b>  | <b>Relevance</b>                          | <b>Comfort vs difficulty</b>    | <b>Impact, demand</b>                  |

Figure 2. The nine experiences as involved in studying, constituting 4x2 ‘loops’ (each reflecting a specific type of motivation), their going together resulting in effectiveness, intrinsic motivation or ‘deep level learning’, as such enabling the student the self-regulation of as many operations as involved in his studying as a process.

This effectiveness has to be psychologically understood as intrinsic motivation, implying that the student - whether or not - is experiencing his studying as rewarding in itself. In the positive case he *reaches ‘deep level learning’* (Marton & Säljö, 1976), as such the essential condition for the development of real expertise. In that case this very same intrinsic motivation integrates up to four ‘different’ but all essential motivational subsystems for his optimally functioning as a human being: **(1) competence** motivation (the effectively going together of ‘self confidence and ‘situational comfort’); **(2) causality** motivation (the effectively going together of ‘activity to control’ and environmental ‘relevance’); **(3) curiosity or creativity** motivation (the effectively going together of ‘intention’ or personal interest and situational ‘demand’) and **(4)**, last but not least, *stability* or self regulation motivation (the effectively going together of ‘agenda’ or time perspective and ‘discipline’).

In the meantime the parallelism between figures 1 (operations involved) and 2 (the complementing study experiences) merits further attention. So, indeed, the process of studying (figure 1) can be regulated by the individual student on the base of his *experiences* (figure 2); so the 'central' experience of effectiveness, as such resulting from the permanent evaluation of the ongoing process, becomes the necessary and sufficient situational knowledge, which enables the student, so becoming a studax or real expert in studying (Janssen, 1996), to be(come) and to remain effective in Higher Education. *'Only' three conditions have to be fulfilled: (a) a sufficient expertise in studying (which can be achieved by everyone who before has 'learned to learn' during secondary education), (b) a sound vocational decision (as such resulting into an internally well coherent matrix of the study experiences to come) and, (c) last but not least, a sufficient ... competence (as described in the preceding paragraph).*

Both last conditions imply each other to a certain degree. Indeed, it is quite probable that the person who is making a sound vocational decision, will detect a potential conflict between his competence and the intellectual requirements as involved in a specific alternative, in due time and will take the consequences quite seriously... He spontaneously will change his planning towards another alternative offering him the highest 'value for money', i.e. the highest guarantee for real effectiveness. So this 'hecatomb' in the transition from Secondary towards Higher Education - at least in Flanders - can definitely be cleared up: *One has to 'organize' the counselling that last year students in Secondary Education need, in such a way, that each of them is capable to take into consideration his strong as well as his weak competencies before, or at least while, making his vocational decision. A careful analysis of that vocational decision - as such it turns out to be a process, during which up to five stages have to be passed through - reveals the necessity to offer these youngsters the right information at the right moment, so enabling them to take autonomously that decision, which allows full effectiveness in what they, by consequence, will undertake.*

### **3.3. An effective vocational decision: the necessary and sufficient condition**

It is quite clear that society as such has to promote an optimal equilibrium between its societal needs and this psychological effectiveness of each of its members. That may motivate its government to take the counselling of its last year students in Secondary Education - it refers to the immediate future these youngsters themselves are planning - as seriously as possible. All kinds of measures can be developed in order to promote that, given the fact that this process of vocational decision psychologically turns out to be an individual process of matrix construction, selecting that *environment* that as a task effectively fits the *person* one is.

The crux is to meet, as effectively as possible, the ‘different’ needs these youngsters have at each of the up to five ‘different’ stages they are going through, in order to attain a firm decision with respect to their future on the immediate as well as on the long run. They vary from **(1)** a good explanation of the implications of the making of a vocational decision in the very first stage of **sensitisation**, to **(2)** the presentation of relevant information about occupations and their development within a modern society during the essential second stage of **horizon enlargement**, **(3)** the opportunity to test themselves with respect to their potentials, entering so their third stage of **self-concept clarification** with respect to each of the alternatives which seemed relevant at the end of the preceding stage and last but not least - after **(4)** the making of their **decision** - the **(5)** Reality testing of what was decided during the very important ‘last’ stage of elaboration. Here the entrance exam for medicine and dentistry, by definition, has to fulfill, as the other ones in the faculties of engineering and in the artistic study domains of Higher Education outside universities already do, an essential condition for the promotion of ‘deep level learning’ or effective studying.

As a member of the Flemish community in Belgium as my country, I hope that my Minister of Education will realize himself in due time also, that the **timing** of this entrance selection is as important as the definition of its content. For that reason I propose to organize it’s first session at Eastern and its second at the end of June. So students who will not be allowed to register in medicine or dentistry, can take *in due time* a new decision that offers them elsewhere that same ‘effectiveness’, which means ‘personal happiness’, they *really* are looking for.

## References

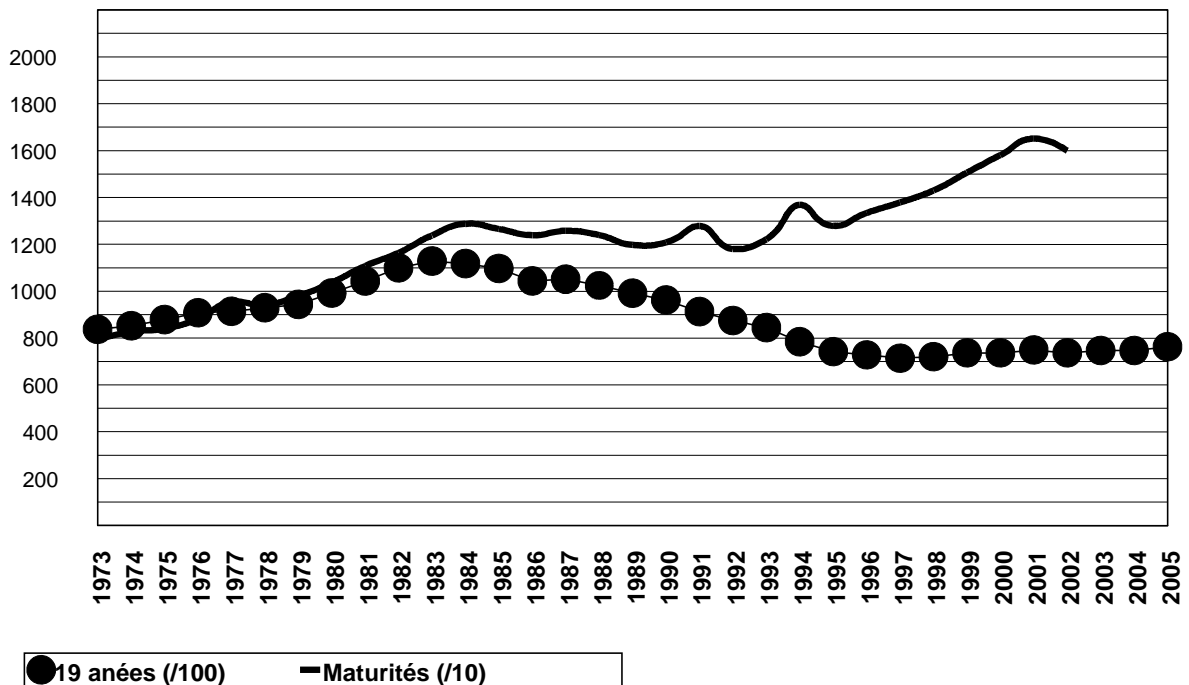
- Ausubel, D. (1963). The psychology of meaningful verbal learning. An introduction to school learning. San Francisco/London: Jossey Bass Inc.
- CgAO (Contactgroep Academisch Onderwijs) (1993). Van secundair naar hoger onderwijs; is er een einde aan die jaarlijkse hecatombe? Leuven: Garant.
- Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.
- Janssen, P.J. (1996). Studeren, doceren en evalueren in studaxologisch perspectief. Leuven: Acco.
- Janssen, P.J. (1996). Studaxology: the expertise students need to be effective in higher education. *Higher Education*, 31, 117-141.
- Janssen, P.J. & De Neve, H. (1988). Studeren en doceren aan het hoger onderwijs; vakmanschap als leeropdracht. Leuven: Acco.
- Marton F. & Säljö, R. (1976). On qualitative differences in learning. 1. Outcome and processes. 2. Out-come as a function of a learner’s conception of the task. *British Journal of Educational Psychology*. 48, 4-11 and 115-127.
- Minnaert A. & Janssen, P.J. (1992). Success and progress in Higher Education: a structural model of studying. *British Journal of Educational Psychology*, 62, 184-192.
- Pask. G. (1976). Conversational techniques in the study and poractice of education. *British Journal of Educational Psychology*, 46, 128-148.

## Le test d'aptitude aux études de médecine en Suisse en tant qu'instrument de limitation de l'accès aux études

**K.-D. Hänsgen**

Un regard réaliste sur les tendances internationales présentées dans ce volume suffit à s'en persuader: il est peu probable que le problème du dépassement des capacités d'admission aux études de médecine en Suisse se résoudra spontanément. Dans tous les pays industrialisés on voit la population aspirer à une meilleure formation et grossir le nombre de personnes ayant accès aux études universitaires. Les études de médecine exercent un attrait particulier, la profession médicale gardant pour diverses raisons un prestige enviable. Cela étant, la plupart des pays ont dû recourir à un *numerus clausus* pour adapter le nombre d'admissions aux capacités de leurs universités.

Quels sont les pronostics? Dans le tableau ci-après figurent les inscriptions en faculté de médecine et en parallèle, par année, le nombre de maturités (divisé par 10) et la population résidante âgée de 19 ans (divisée par 100).



Les chiffres des maturités à partir de 1996 sont tirés de l'estimation du CSDOC (Centre suisse de documentation en matière d'enseignement et d'éducation). Malgré un taux de population en recul ces dernières années, et qui stagnera à l'avenir, le nombre des maturités est toujours en hausse. On peut donc admettre que la pression se maintiendra sur les universités. En médecine, selon la tendance propre à cette discipline, le nombre des nouveaux étudiants continue d'augmenter.

Le degré d'adhésion à un *numerus clausus* (NC) en Suisse sera d'autant plus élevé que le critère retenu pour la "régulation" des admissions sera juste et scientifiquement éprouvé. Mais il faut se décider: Introduire un NC veut dire sélectionner avant les études. La décision politique serait finalement vaine si en dépit du NC presque tous les candidats devaient néanmoins être admis. Le référendum dans le canton de Berne en 1995 a clairement démontré que les 2/3 environ des votants sont favorables à la création de bases légales pour un NC.

Selon quels critères les places d'études pourraient-elles être attribuées? Une sélection avant le début des études exige des critères acceptables, objectifs et équitables, qui résistent également à une contestation juridique. Le tableau ci-après résume les avantages et les inconvénients de quatre critères possibles:

|   | <b>Maturité/ moyenne de notes</b>   | <b>Entretien d'évaluation</b>  | <b>Stage pratique</b>   | <b>Test d'aptitude</b>  |
|---|---|--|---|---|
| <b>Egalité de traitement/ comparabilité</b>                     | condition non remplie<br>notation plus ou moins sévère avec des critères différents selon les cantons et les écoles | une comparaison entre groupes d'entretien nécessite une standardisation de ces entretiens et une formation des experts chargés de l'évaluation                 | à condition que les performances requises et la qualité de l'assistance soient comparables  | entièrement réalisées par la standardisation de l'exécution                             |
| <b>Admission possible en fonction de la capacité d'accueil?</b> | Oui   | Oui  | la concordance ne serait que hasard   | Oui   |
| <b>charge supplémentaire problèmes</b>                          | aucune<br><br>problème: répercussions possibles sur les principes de notation des écoles                            | 1 rendez-vous par candidat, formation des experts,<br><br>3 experts à 30 min. avec évaluation (1800*3*0,5 h) = 2350 journées = 10 ans de travail (irréaliste!) | perte évent. d'une année pour les candidats<br><br>capacité de soins qualifiés à trouver<br><br>recherche de l'équipement approprié et disponible | 1 jour de session, directeurs de la session de test<br><br>l'évaluation est centralisée |
| <b>Adéquation (études de validité)</b>                          | n'entre pas en ligne de compte en raison des différences de notation  | les prévisions de succès dans les études sont moins pertinentes que celles résultant du test   | "intimidation" - manque d'expérience quant à l'efficacité   | le plus favorable<br><br>la possibilité d'être appliqué en Suisse a été vérifiée        |
| <b>reproductibilité</b>   | non   | oui  | pas nécessaire  | oui   |

L'utilisation des **notes de maturité** en tant que critère d'accès aurait surtout deux très graves inconvénients: d'abord, la quasi impossibilité de combler les différences entre cantons ou groupes linguistiques et à l'intérieur de ceux-ci. On rencontre des différences par exemple dans les taux de maturité (proportion de maturités délivrées pour le nombre d'étudiants d'une classe en fin de scolarité), dans les critères d'attribution des notes et la sévérité dans la notation; second inconvénient: les répercussions possibles sur le système de notation lui-même si les enseignants tiennent compte du rôle joué par les notes dans l'admission aux études.

L'idée de soumettre à des **entretiens d'évaluation** tous les candidats d'une classe conduit à l'absurde: ce serait une charge annuelle équivalant en gros à 10 années de travail. Ce système a toujours ses tenants dans la discussion parce qu'on en attend une preuve de la motivation et des compétences sociales des candidats et que l'on voudrait davantage vérifier au cours de ces entretiens le talent, les dispositions professionnelles des candidats. Toutefois, vu la faible valeur pronostique reconnue à ce critère parmi les autres pour le succès dans les études (cf. 18e rapport de travail du Centre pour le développement de tests et le diagnostic, Trost, 1995), on ne peut l'utiliser sans retenue et il serait faux de faire dépendre **généralement** du résultat de ces entretiens l'admission aux études de médecine. Les compétences sociales doivent se développer durant la formation. Les entretiens d'évaluation apporteraient plutôt une différenciation dans les signes d'inaptitude aux études de médecine. On pourrait envisager d'y soumettre une partie des candidats. Nous reviendrons sur ce point.

La principale objection au **stage pratique** réside dans la forte dépense qu'il suppose en infrastructure de soins. L'expérience qui a été faite dans le canton de Zurich montre que cette charge supplémentaire pour un système de santé qui lutte pour l'efficacité ne peut être consentie sans réserve et que la mise à disposition de places en suffisance n'est pas réalisable. Mais l'inconvénient majeur du stage pratique, c'est son incapacité à faire concorder le nombre des admissions et des places d'étude. Le stage peut être "trop dissuasif" ou alors ouvrir la porte de l'université à de trop nombreux candidats qui l'auront accompli avec succès. C'est alors le hasard qui déciderait.

### **Le test en tant qu'"études à l'essai"**

Au regard de la législation des cantons, **l'aptitude** aux études apparaît comme l'unique critère d'admission qui soit à la fois réalisable et acceptable. C'est aussi la garantie que selon toute vraisemblance les candidats admis iront au bout de leurs études. En limitant les admissions sans tenir compte de l'aptitude, on court le risque d'un taux d'abandons trop élevé - les étudiants admis ne satisfaisant pas aux exigences, faute des capacités de base nécessaires - et même d'une sous-occupation des capacités d'accueil.

Si l'on élimine la note de maturité et l'entretien d'évaluation en tant que méthodes d'application générale, il ne reste que les tests comme procédés valables pour analyser l'aptitude aux études. Les tests psychologiques sont des procédés scientifiques d'examen visant à renseigner sur les aptitudes ou les comportements des personnes qui y sont soumises. Ils donnent des indications sur les qualités spécifiques de la personne examinée, pour servir de base à des décisions pouvant être de grande portée. C'est dire que ces indications doivent être solidement fondées et qu'avant d'en tirer argument pour décider par exemple d'une admission à un curriculum ou de l'aptitude à une activité déterminée, il y a lieu de les soumettre à un contrôle scientifique parfaitement défini.

Pour satisfaire aux exigences scientifiques, un test psychologique doit remplir plusieurs critères de qualité. Objectivité d'abord: il doit apporter des résultats détachés de la personnalité de l'examineur. Afin d'assurer l'égalité des chances entre candidats, il est indispensable que le test soit exécuté et ses résultats évalués dans des conditions standardisées. Si l'on attend du test des prévisions sur la réussite d'un candidat dans ses études, il faut que les résultats du test soient effectivement en rapport avec cette réussite. C'est ce que l'on appelle en psychologie le contrôle de validité.

Mais le test doit aussi être fiable, c.-à-d. que les problèmes posés dans chaque sous-test doivent mesurer la même caractéristique et les sous-tests mesurer dans leur totalité la même aptitude.

Un **test d'intelligence** vise à détecter des aptitudes d'ordre très général, sans référence à une activité précise. Il dira par exemple si les connaissances de la personne soumise au test et son aptitude à l'action correspondent à son âge. A l'aide de ces résultats on déterminera par exemple sa maturité pour l'école. Le test d'intelligence mesure donc un savoir global, étendu, ou des aptitudes très générales.

Un **test d'aptitude**, par contre, livre des résultats fiables sur la capacité de satisfaire à une exigence concrète, d'exercer une activité spéciale. Il sert de base à un pronostic sur la capacité du candidat à réussir dans une activité; il permet d'évaluer son habilité à surmonter les difficultés qui pourraient se présenter.

Le test d'aptitude aux études de médecine (TMS) développé en Allemagne entre dans cette dernière catégorie de tests. Conçu en fonction des études de médecine, il livre un pronostic de probabilité pour le participant au test à remplir les exigences des études de médecine de base.

Le contenu du test devrait être en rapport très étroit avec les études de médecine. C'est pourquoi les questions ont trait aux études de base en médecine. Pour plusieurs raisons, le test se limite à l'appréciation des performances; les compétences sociales et les aptitudes à communiquer, au demeurant difficiles à cerner dans un test, n'y sont pas prises en compte. Faute de ressources suffisantes, des entretiens d'aptitude standardisés n'étaient pas réalisables. Du

reste, ces compétences se développent plutôt entre la maturité fédérale et la fin des études, d'où la faible valeur prédictive des indications qu'on aurait pu recueillir dans ces entretiens.

Le test d'aptitude planifié comprend neuf sous-tests, c.-à-d. neuf groupes de problèmes du même type, que les participants doivent résoudre en 5 heures environ: cinq sous-tests le matin et quatre l'après-midi, avec une heure de pause à midi.

Chaque sous-test est précédé d'une brève introduction expliquant la visée de cette partie du test. Toutes les épreuves d'un sous-test sont construites selon le même principe. A chaque problème - sauf dans le test de concentration -, cinq réponses sont proposées, dont une seule est correcte. Les participants disposent d'un temps limité pour chaque sous-test.



Tableau 1: Structure et déroulement du test pour les études de médecine.

Source: Trost, Günter (Hrsg.) (1994): Test für Medizinische Studiengänge (TMS). Studien zur Evaluation. 18. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.

| Désignation des sous-tests  | Aptitudes mesurées   | Nombre des items*              | Temps de traitement (en min.) |
|---|--|--------------------------------|-------------------------------|
| Reconnaître une partie d'une figure                                 | différenciation visuelle   | 24 (20)                        | 22                            |
| Compréhension de questions médicales et des sciences naturelles     | capacité de compréhension et de représentation de problèmes relevant de la médecine et des sciences naturelles | 24 (20)                        | 60                            |
| Figures tubulaires  | représentation spatiale  | 24 (20)                        | 15                            |
| Problèmes quantitatifs et formels                                   | résolution de problèmes quantitatifs relevant du contexte des sciences naturelles et de la médecine            | 24 (20)                        | 60                            |
| Aptitude à travailler de manière concentrée et soigneuse            | concentration et attention   | 1200 signes                    | 8                             |
| <b>P a u s e</b> 60 minutes   |  |                                |                               |
| Phase de mémorisation d'éléments (examen des aptitudes à mémoriser) |  |                                |                               |
| Retenir des figures   |  | 20 unités                      | 4                             |
| Retenir des faits   |  | 15 unités                      | 6                             |
| Compréhension de textes   | compréhension et interprétation de textes médicaux et des sciences naturelles                                  | 24 (18)                        | 60                            |
| Tests de la mémoire   |  |                                |                               |
| Retenir des figures   | mémorisation de figures  | 20 (20)                        | 5                             |
| Retenir des faits   | mémorisation verbale   | 20 (20)                        | 7                             |
| Diagrammes et tableaux  | interprétation de diagrammes et tabelles   | 24 (20)                        | 60                            |
| Test entier   |  | 204 (178)                      | env. 5 heures                 |
| Durée totale (pauses comprises)                                     |  | de 8.45 h env. à 16.00 h. env. |                               |

\* Entre parenthèses, nous citons les épreuves évaluées; les autres questions sont des questions supplémentaires qui seront utilisées pour d'autres tests, et qui doivent être évaluées dans des conditions réelles pour l'un des tests suivants.

## Le test est-il assez spécifique pour la Suisse?

Le fait d'adapter le test allemand offre d'abord l'avantage de ne pas devoir partir de zéro mais de disposer sans détours hasardeux d'un test apte à pronostiquer le succès ou l'échec dans les études. Cela n'a pas empêché de prendre en considération les particularités helvétiques. En raison de l'urgence d'une solution dans notre pays, il fallait avoir très vite un test de bonne qualité et le test allemand est bien coté en comparaison internationale. Une conception entièrement suisse aurait pris au moins 2 ans. En effet, avant d'être incluse dans le test, chaque question doit être mise à l'épreuve dans des conditions réelles. D'après les informations des concepteurs du test, 1/3 seulement des questions élaborées par les meilleurs experts répondent vraiment aux critères définis pour leur application dans le test. Les experts eux-mêmes ne peuvent évaluer qu'approximativement le degré de difficulté. Seul le passage des questions au banc d'essai dans des conditions réelles - elles sont insérées dans un test original - garantit que les prestations des candidats pourront être différenciées de manière sûre et optimale.

En juin 1995, une étude empirique a été réalisée à Fribourg: le test sous une forme préliminaire a été expérimenté dans des conditions réelles sur un petit échantillon d'étudiants (Hofer, Ruefli, Hänsgen 1995). Afin de vérifier les critères de qualité psychologique du test et son équité, un échantillon de lycéens et lycéennes - 54 de langue allemande et 126 de langue française - ont passé le test au Collège Sainte-Croix à Fribourg (Suisse).

A noter que ce "galop d'essai" ne s'est pas déroulé dans les conditions usuelles d'une compétition. La seule "motivation" qui a pu être donnée aux participants est une appréciation individuelle de leurs prestations.

En dépit de ce défaut de motivation, le test a montré dans cet essai, à peu de chose près, les **mêmes critères de qualité** et la même fiabilité qu'en Allemagne. En moyenne, la moitié environ (48%) des problèmes ont été correctement résolus et les valeurs s'étalent sur une **largeur optimale pour la différenciation des performances**.

Les deux **groupes linguistiques** ont obtenu d'aussi bons résultats et l'on peut conclure à un haut degré d'équivalence des deux formes linguistiques.

En ce qui concerne la valeur prédictive des résultats du test pour le succès aux examens propédeutiques, on a constaté dans le groupe francophone que les moyennes obtenues au test par les étudiants et étudiantes en classe de maturité de type D sont significativement inférieures aux moyennes des types A, B et C. Cela coïnciderait avec le résultat des examens de médecine à Berne, où le taux d'échec au premier tour du premier propédeutique est de 44 à 50% pour les

détenteurs de maturités A à C et de 77% pour les maturités de type D (Hofer, 1992).

Globalement, les résultats de cet essai prouvent que **le test serait tout aussi bien utilisable en Suisse en tant que critère d'admission équitable et fiable**. Ils n'ont pas confirmé de différences de performances entre sexes, groupes linguistiques ou même en comparaison avec l'Allemagne, pays d'origine du test.

### **Le test offre-t-il la garantie d'un traitement équitable (égalité des chances)?**

Le test se distingue des autres critères par le fait qu'il réunit les conditions d'un traitement équitable de tous les candidats. C'est là un avantage particulièrement important pour la Suisse.

Le test est organisé simultanément à tous les endroits prévus, dans des conditions similaires. Grâce à la brochure d'information qui a été publiée, les candidats ont la possibilité de s'y préparer tout à loisir. La passation du test est standardisée; les conditions du local, le temps imparti et l'information orale donnée aux participants sont exactement prescrits. Le personnel chargé de l'exécution reçoit au préalable une large instruction. L'évaluation est objective et aucun doute ne peut subsister sur ce qui est juste ou faux.

Aucun facteur subjectif n'entre dans l'appréciation, ce qui serait difficile à éviter, par exemple, dans un entretien.

Dans l'application d'un test de ce genre, le risque de désavantager des jeunes issus de couches moins favorisées est aussi relativisé. En Allemagne, depuis des années, on enregistre l'appartenance sociale des participants au TMS. Il est réaliste de dire qu'un assez bon équilibre a pu être maintenu entre les candidats à l'admission et ceux effectivement retenus, et ce à travers toutes les couches sociales. Une raison à cela pourrait être que d'éventuelles inégalités à l'école sont compensées par le fait que le test ne mesure pas uniquement la quantité des connaissances acquises. Les étudiants trouvent là, en réalité, une "nouvelle chance".

### **L'égalité des chances vaut-elle pour les trois groupes linguistiques?**

Il faut être particulièrement attentif à l'égalité des chances entre groupes linguistiques. A cet égard se pose aussi tout naturellement la question de l'équivalence des trois versions du test quant au degré de difficulté. Pour les traductions, une procédure correspondant aux normes internationales a été appliquée; elle comprend les phases suivantes:

- première traduction par un traducteur spécialisé dans le domaine scientifique et dont la langue maternelle est la langue cible;
- examen de cette traduction par des experts bilingues;
- re-traduction en langue allemande par un second traducteur indépendant du premier et comparaison de sa version avec le texte d'origine;
- conférence de tous les participants pour débattre des divergences et trouver la formulation définitive des passages discutés.

Il en résulte trois formes linguistiques équivalentes qui, vues sous l'angle scientifique, garantissent un haut degré d'équité. Que le test soit passé en allemand, en français ou en italien, le résultat sera le même.

### **Les femmes sont-elles désavantagées par le test?**

Il est juste d'exiger qu'hommes et femmes aient les mêmes chances d'accès à l'université et que le test ne mette pas en péril les progrès réalisés. Si des craintes dans ce sens ont été exprimées de divers côtés, un examen attentif de la question montre qu'elles ne sont pas tout à fait justifiées et surtout que le test ne doit créer aucun préjudice aux femmes.

Dans l'évaluation du test, on parle toujours des résultats du TMS allemand. C'est un fait reconnu: statistiquement, la moyenne des notes d'examen de médecine obtenue par les femmes est dans ce pays inférieure à celle des hommes. De nombreuses analyses l'ont confirmé année après année. Il est vrai aussi que dans le test également, les femmes recueillent des notes inférieures de 2 points à celles des hommes (pour une valeur moyenne de 100 et un écart standard de 10). En termes relatifs la différence selon le sexe dans les notes d'examen est encore plus importante que dans le test. On pouvait s'y attendre si l'on veut voir dans le test des "études à l'essai" qui prédisent effectivement la performance dans les études universitaires. C'est même logique dans la mesure où le test doit refléter la réalité toute nue.

Tout dépend de la manière dont on traite les faits. Que cette différence ait été constatée ne signifie pas forcément qu'elle sera retenue telle quelle dans les critères d'admission. Si une décision politique l'autorise, il est possible d'appliquer un procédé de correction statistique qui aurait pour effet de compenser la différence de valeur moyenne entre les deux sexes. Les femmes auraient ainsi en moyenne les mêmes résultats de test que les hommes. Les quotas d'admission correspondraient ainsi pour les deux sexes exactement à ceux des candidatures.

L'application de cette procédure suppose toutefois, outre une décision politique favorable, la preuve que ces différences existent aussi en Suisse. L'hypothèse tirée des résultats allemands, selon laquelle les **femmes atteignent de moindres**

**performances en général** que les hommes n'a **pas été confirmée** lors du test d'essai organisé en Suisse. Dans le groupe germanophone les femmes ont même obtenu de meilleurs résultats que les hommes et dans le groupe francophone on n'a pas relevé de différence significative. Il faudrait donc attendre ce que donne sur ce point le test définitif.

### **Des cours de préparation sont-ils nécessaires?**

C'est un fait établi et un avantage particulier du test allemand: il est superflu de s'y entraîner par des cours s'ajoutant à la préparation recommandée. La preuve en a été donnée par plusieurs enquêtes de l'Institut de recherche sur les tests et les aptitudes, à Bonn, ainsi que par une étude indépendante. La préparation recommandée se limite à la lecture attentive d'une brochure d'information sur le test et à l'étude d'une version originale du test qui a été publiée. Des cours ou livres d'exercice supplémentaires proposés par des maisons commerciales n'apportent aucune amélioration notable des performances (voir exposé Hofer et Hänsgen dans ce cahier).

### **Un mot sur le coût du test**

L'argument du coût invoqué contre une "bureaucratie de l'admission aux études" est un raisonnement à court terme: l'introduction du test en Suisse ne nécessite que des structures relativement simples. En tablant sur une dépense de 850.000 francs pour le test, cela fait l'équivalent de 17 places d'étude si l'on estime à 50.000 francs le coût d'une place pour les études de base en médecine. Autrement dit, 10% environ des fonds économisés en tout par une régulation initiale des admissions sur le plan suisse suffiraient pour financer le test. Bon à prendre, ce "rendement" de 90% pourrait encore augmenter des fonds qui se perdent aujourd'hui dans les études interrompues dans les deux premières années (60%) et dont une large part ne serait plus nécessaire.

### **Conclusions et évolutions possibles**

Lorsqu'il s'agit de décider **sur la base de quels critères** un numerus clausus peut être **réalisé** en Suisse, **la préférence va, comme avant, à un test d'aptitude** - de la même manière qu'aux Etats-Unis et au Canada par exemple, en Suède, en Tchéquie, en Finlande, en Israël, au Japon, ou en tant que test de connaissances standardisé dans 17 autres pays européens. En raison des particularités suisses (notes de maturité guère comparables entre les cantons et les régions linguistiques; impossibilité - pour raisons financières - d'organiser des entretiens d'évaluation pour tous les candidats, difficultés de réaliser des stages pré-universitaires), un tel test d'aptitude - adapté dans son application aux

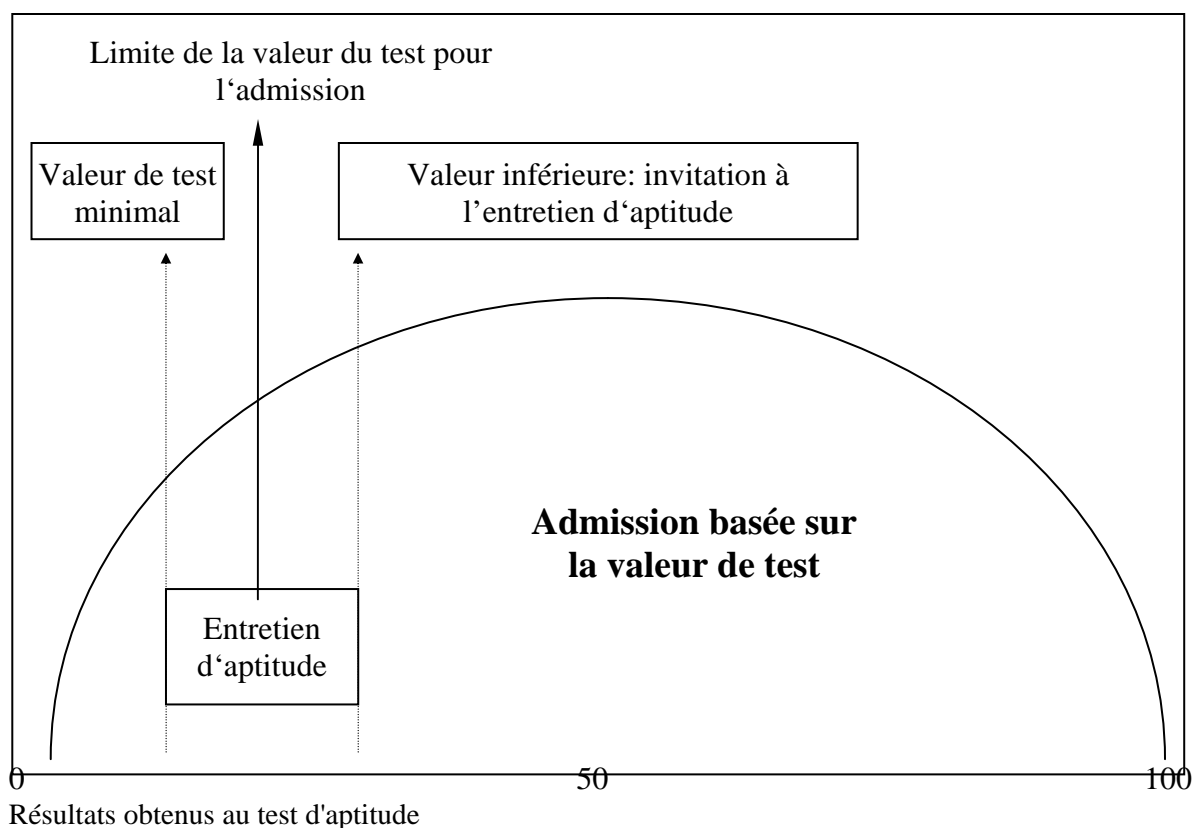
conditions suisses - reste **le critère le plus juste et le mieux garanti scientifiquement.**

Le test allemand TMS retenu comme base en Suisse a été construit et élaboré avec beaucoup de soins; il se distingue en comparaison internationale par sa haute valeur prédictive de réussite ou d'échec dans les études. Le coût de **réutilisation** en Suisse d'un test allemand existant - pour une même qualité prévisionnelle - se monte à moins d'un dixième des frais engagés en Allemagne. C'est sans doute un avantage du concept choisi en Suisse. A noter que toute forme de test doit être renouvelée chaque année afin que les problèmes posés ne soient pas connus des candidats.

Si un NC est nécessaire en Suisse, on devra examiner à long terme (1) si **d'autres tests de qualité égale** peuvent être conçus à **moindres frais** et (2) **si des méthodes complémentaires sont indiquées.** Peut-être faudra-t-il compléter l'analyse de l'aptitude aux études par d'autres signes d'aptitude à l'exercice de la profession; en pareil cas on pourrait songer à des entretiens d'évaluation standardisés à l'intention d'une partie des candidats, pour déterminer leur motivation et leur compétence sociale. De tels entretiens sont toutefois liés à d'assez hautes exigences, par exemple la comparabilité des résultats entre divers groupes d'interviewers et la capacité réellement assurée et scientifiquement établie de réguler les admissions (par exemple par une valeur de point analogue à celle du test).

A l'heure des économies, il serait peu réaliste d'organiser sur le plan suisse quelque 2000 entretiens d'aptitude. C'est pourquoi la **combinaison d'un test obligatoire et d'entretiens** pour un nombre limité de candidats mériterait réflexion. On pourrait de cette manière attribuer la grande **majorité des places d'études d'après la performance réalisée dans le test**, c'est-à-dire aux candidats dont les prestations attestent le mieux l'aptitude à répondre aux exigences des études de médecine.

Seraient invités à un entretien les candidats dont le **score au test** est insuffisant pour obtenir **une place d'études.** Ou bien l'on convoque tous ces candidats, ou bien l'on fixe là aussi un critère, soit un **score minimum.** On toucherait ainsi de **manière très ciblée des personnes** ayant obtenu au test des résultats se situant dans un champ défini autour de la "valeur limite" des admissions possibles par an. En procédant de cette manière, on arrive à prévoir très exactement la capacité annuelle nécessaire pour les entretiens d'aptitude, comme le montre le schéma ci-après.



Les places d'études restantes seraient attribuées aux candidats entendus qui, vu le résultat de l'entretien, sont le plus aptes aux études. Les prestations réalisées dans le test pourraient être prises en considération sous une forme pondérée; il serait ainsi possible de compenser de mauvaises prestations au test par de meilleures compétences sociales.

Les entretiens d'aptitude doivent être standardisés. Les thèmes abordés sont fixés d'avance. 3 experts participent à chaque entretien, au terme duquel ils donnent leur impression, fondée sur divers critères, dans une feuille d'évaluation standard. Celle-ci aboutit de même à une valeur globale du point qui rend comparables les candidats répartis dans les groupes (addition des points attribués par les 3 experts). Des différences systématiques (émanant des experts) entre les groupes pourraient être compensées statistiquement.

Le coût supplémentaire pour le travail d'élaboration du concept et de la documentation par le Centre pour le développement de tests et le diagnostic serait probablement minime, le surcroît de frais de personnel apparaissant peu important.

*Nous préconisons néanmoins d'utiliser d'abord comme critère le test d'aptitude. Il a été scientifiquement contrôlé, il est utilisable en Suisse et il existe en trois formes linguistiques équivalentes. Le système d'attribution des places d'études qu'il permet de réaliser est correct et - on a pu le constater à l'occasion de cette*

*session - il est reconnu sur le plan international. La recherche de perspectives et d'extensions nécessitant du travail et des préparatifs supplémentaires ne devrait pas empêcher que quelque chose se fasse aujourd'hui - car il y a quelque chose à faire!*

## **Bibliographie**

- Hänsgen, K.-D., Hofer, R., Ruefli, D. (Hrsg.) (1996) Eignungsdiagnostik und Medizinstudium. Tagungsband. Berichte des ZTD Band 2 Univ. Fribourg.
- Hänsgen, K.-D., Hofer, R., Ruefli, D. (1996). Un test d'aptitudes aux études de médecine est-il faisable en Suisse? Bulletin des médecins suisses, 7, S. 267 - 274.
- Hänsgen, K.-D., Hofer, R., Ruefli, D. (1995). Der Eignungstest für das Medizinstudium in der Schweiz. Schweizerische Ärztezeitung, 37, S. 1476 - 1496
- Hofer, R., Ruefli, D., Hänsgen, K.-D. (1996). Der Eignungstest für das Medizinstudium in der Schweiz. Ein Probelauf. Berichte des ZTD Band 1 Univ. Fribourg
- Trost, Günter (Hrsg.) (1978-1994): Test für Medizinische Studiengänge. 1. - 18. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Zentrum für Testentwicklung (1995). Il test attitudinale per lo studio della medicina (Adattamento italiano). Göttingen: Hogrefe
- Zentrum für Testentwicklung (1995). Le test d'aptitudes pour les études de médecine (Adaptation française). Göttingen: Hogrefe



## **Test d'aptitude aux études de médecine en Suisse: dans quelle mesure peut-on s'y entraîner?**

**R. Hofer & K.-D. Hänsgen**

Les velléités d'introduire en Suisse un test d'aptitude en tant que procédure de sélection pour l'accès aux études de médecine soulèvent notamment cette question: dans quelle mesure est-il possible de s'entraîner aux épreuves du test. La passation du test ne devrait pas causer aux candidats une dépense supplémentaire déraisonnable en temps ou en argent. Si la fréquentation de cours d'entraînement améliorerait sensiblement les chances d'admission, l'admission dépendrait moins de l'aptitude aux études. Ce qui compterait surtout, c'est d'être prêt à consacrer les ressources nécessaires à un tel entraînement et d'en avoir les moyens.

Tel qu'il est conçu, le test allemand pour l'accès aux études de médecine (TMS) met l'accent sur l'aptitude du candidat à élaborer de nouvelles solutions et à répondre à de nouvelles exigences, et non pas sur le contrôle des connaissances assimilées. Tous les éléments nécessaires pour résoudre la question posée se trouvent dans le libellé même de la question. La préparation au test se fait au moyen d'une brochure contenant des exemples de devoirs et des exercices conseillés. Elle a été conçue en tenant compte de l'influence que peuvent avoir sur la performance la répétition de certaines tâches et le fait d'être familiarisé avec le genre de questions abordées dans le test. Aux dires des concepteurs du test, la brochure dispense ce qui est nécessaire à la préparation et rien de plus. On peut se demander dès lors si cette préparation "standardisée" suffit et si d'autres mesures n'apporteraient pas des incidences supplémentaires. Le cas échéant, la discussion avec les prestataires de programmes d'entraînement peut et doit être intense et offensive. On sait par l'expérience faite en Allemagne que l'argumentation de ces offres est souvent axée sur le recrutement de participants (payants) plutôt que sur la réalité.

Dès l'introduction en Suisse du test d'aptitude aux études de médecine, on verra probablement se développer ici l'offre de livres et de cours d'entraînement au test, promettant une nette montée de la performance et donc des chances d'admission aux études de médecine ou l'échec à qui n'a pas suivi ces cours. Finalement, ce genre de publicité exploite l'angoisse ressentie par les candidats, sentiment bien compréhensible et jusqu'à un certain point inévitable dans une situation inconnue jusqu'ici. Avec la première introduction d'un test en Suisse, des béotiens en la matière tombent sur des "formateurs" opérant en Allemagne, dont les slogans publicitaires sont bien rodés.

Qu'une motivation renforcée augmente les chances d'être admis aux études, si elle a été mise à l'épreuve par un grand travail de préparation au test, n'aurait rien de néfaste en soi. Mais si le pouvoir d'achat devait influencer l'intensité de cette préparation, les mieux lotis pourraient obtenir sur les autres un avantage injustifié. C'est pourquoi en Allemagne aussi, la question de l'influence de l'entraînement s'est rapidement posée, avec la nécessité de la contrôler scientifiquement (en résumé, voir Deter 1982, p. 12 s.).

On trouve déjà chez Throndike (1919, cité dans Deter 1982, p. 14) de premières conjectures au sujet de l'influence de l'entraînement sur la force d'expression des tests d'intelligence: "In proportion as such tests are used for promotion in schools, entrance examinations for colleges or professional schools, employment examinations in business and industry, or any purpose where a high score is to the advantage of the person tested, the danger from deliberate coaching become very grave."

Depuis lors, cette question occupe les esprits chaque fois qu'est organisé un test dont les résultats servent à une sélection. Ainsi en fut-il par exemple aux Etats-Unis lorsque dans les années vingt le premier test d'intelligence pour des enquêtes de groupes "Army Alpha" fut mis à disposition pour être utilisé dans le civil, ou à la fin des années quarante en Grande-Bretagne, quand fut introduit l'examen "Even-plus" pour l'admission dans les lycées britanniques (cf. Deter, 1982). Il en alla de même à la fin des années 50 aux Etats-Unis lorsque les "Scholastic Aptitude Tests" (SAT) gagnèrent en importance pour l'admission dans les collèges ou encore à la fin des années 70 en Allemagne, au moment où fut projeté le premier lancement à l'essai du test pour les études de médecine (TMS).

Le TMS comprend neuf sous-tests, c.-à-d. neuf groupes de questions du même type. Les contenus du test sont en rapport très étroit avec les études de médecine (cf. Trost et al. 1976-1996). Sont examinées les aptitudes de différenciation visuelle (sous-test "attribuer des figures à un modèle"), la compréhension de problèmes ("compréhension de problèmes relevant de la médecine et des sciences naturelles"), la représentation spatiale ("figures tubulaires"), le maniement des chiffres, grandeurs, unités et formules ("problèmes de quantités et de formules"), la concentration au travail ("travail concentré et soigneux"), la mémorisation ("retenir des figures", "retenir des faits"), la compréhension et l'interprétation de textes ("compréhension de textes") et l'analyse de diagrammes et tableaux ("diagrammes et tableaux").

Il est certain qu'une approche préalable du test apporte des avantages. Le seul fait de s'être familiarisé avec les instructions fait gagner du temps au moment critique. Des indications données sur certaines tâches simplifient la recherche de la solution. Pour certains sous-tests, la concentration par exemple, on améliore sa performance par l'exercice.

En Allemagne, suivant cette idée, on a remis gratuitement à toutes les personnes concernées une brochure sur le test qui rassemble des mesures de préparation standardisées, d'une ampleur raisonnable et, de l'avis des concepteurs du test, suffisante. La préparation standardisée dans cette brochure comprend:

- la connaissance du déroulement du test (pauses, repas, matériel nécessaire)
- la connaissance des divers groupes de tâches à l'aide d'exemples et conseils pour la solution, avec possibilité de résoudre certains problèmes types; à chaque sous-test du TMS sont joints les instructions originales ainsi que plusieurs exemples de problèmes avec solutions et explications
- l'explication précise du fonctionnement d'un test: comment sont calculées les notes, quelles stratégies sont recommandées
- conseil de s'exercer plusieurs fois au test de concentration
- le conseil de faire une fois pour soi-même le test (une des deux versions originales qui ont été publiées) en temps réel

Ces conseils et indications sont constamment remaniés et actualisés d'après le résultat d'une évaluation permanente de l'application du test en Allemagne.

Les cours d'entraînement au test proposés par des tiers sont très diversement organisés, d'où la difficulté d'en comparer les effets (cf. Deter 1982, p. 45 ss). La durée des cours va d'une heure à peine, à des séances s'étalant sur une année entière. Quant au contenu, les offres vont de la simple exécution d'exercices à des cours intensifs ciblés, où des stratégies sont exposées et exercées afin de préparer les candidats le mieux possible. Non seulement les cours d'entraînement, mais également les études d'évaluation sont organisées de manière très diversifiée (avec/sans groupe de contrôle, avec/sans test préalable, participants venant de populations diverses, etc.). A part l'étude de Deter (1980, 1982), on n'a pas de données sur les possibilités de préparation pour les groupes de contrôle qui n'ont pas participé aux cours d'entraînement.

Quels sont les motifs qui poussent les candidats à suivre un cours d'entraînement? Selon leur propre estimation, la majorité des participants aux cours se jugent inaptes à travailler seuls et de manière concentrée, sans instruction et des heures durant, sur le matériel de préparation (cf. Mispelkamp 1987). Allalouf (1996) constate qu'ils veulent avant tout se familiariser avec les instructions, les items, les temps impartis et les réponses du test ainsi que recueillir sur le test le plus de documentation possible. Ils espèrent augmenter leur "test-wisness", autrement dit s'exercer à appliquer dans la passation du test des stratégies et des techniques optimales dans le but d'obtenir des notes plus élevées, sans égard par rapport au contenu des questions posées dans le test. Lors de l'évaluation de sept procès-verbaux d'observation où sont décrits la grande majorité des cours d'entraînement proposés en Allemagne (Mispelkamp 1987), les stratégies préconisées par les instituts de formation ont été examinées.

Ces stratégies (répartition du temps, erreurs à éviter, stratégie du décryptage, solution par déductions) ne vont guère au-delà des indications et conseils que donne dans sa brochure le Centre d'attribution des places d'études (ZVS). Apprendre ces stratégies avant de participer au test sera certes très utile (cf. Deter 1982, p. 24 ss). Il est clairement démontré que les candidats ayant déjà fait précédemment l'expérience de questionnaires avec réponses à choix multiple entrent plus facilement dans le système du test que leurs camarades confrontés pour la première fois à une méthode de test standardisée (Trost et al. 1977-1996).

A chaque séance de passation du test, l'étude des effets induits par les divers moyens de préparation fait partie de l'analyse des rapports entre le score obtenu au test et la spécificité des participants (Trost et al. 1977-1996). D'après les réponses données à la question "Vous est-il déjà arrivé de travailler sur des problèmes analogues à ceux qui se présentent dans le test pour l'admission aux études de médecine?", les candidats peuvent se répartir en cinq groupes: (A) ceux qui n'ont jamais abordé de tels problèmes, (B) ceux qui ne se sont préparés qu'à l'aide de la brochure, (C) ceux qui ont utilisé la brochure et l'une des versions du test qui ont été publiées, (D) ceux qui se sont préparés à l'aide de la brochure, d'une version originale du test et d'un livre d'exercice; enfin (E) ceux qui ont étayé leur préparation, entre autres moyens, sur un cours ou un séminaire d'entraînement.

*Influence des moyens de préparation sur les résultats obtenus au test; moyenne des 6 séances de passation du test des années 1982-84.*

| Groupe | préparation   | moyenne* |
|--------|---|----------|
| (A)    | aucune  | 94,7     |
| (B)    | brochure  | 99,4     |
| (C)    | brochure + version originale publiée                        | 104,4    |
| (D)    | brochure + version originale publiée + livre d'exercice     | 105,8    |
| (E)    | brochure + version originale publiée + cours d'entraînement | 104,4    |

\*valeur moyenne 100, écart standard 10

Si l'on résume les résultats des six séances des années 1982 à 1984, on constate que les candidats classés dans le groupe A ont obtenu en moyenne, et de loin, le plus faible nombre de points, soit 94,7 sur une valeur moyenne de 100 et un écart standard de 10 points. La plus forte moyenne (105,8 points) a été réalisée par les candidats qui, selon leurs propres déclarations, se sont préparés comme indiqué pour le groupe D. Les groupes C et E se classent également très bien avec leur moyenne de 104,4 points chacun, tandis que les candidats qui n'ont utilisé "que" la brochure sur le test sont en position médiane avec 99,4 points.

Les groupes C et E obtenant les mêmes résultats, on pourrait en déduire que les cours d'entraînement n'offrent aucun avantage sur la préparation standard. Dans l'appréciation de ces résultats il faut néanmoins se rappeler que les groupes ne sont sûrement pas égaux au départ mais que des différences de potentiel et de motivation déterminent plutôt le choix de la préparation.

Deter (1980, 1982) a mené une expérience dans laquelle fut intégré un groupe qui ne s'était préparé qu'au moyen de la brochure sur le test remise gratuitement. Selon le plan de cette étude les participants furent répartis en cinq groupes dont chacun se prépara différemment au test subséquent. Tandis que le groupe I (n= 96) participait au test préliminaire - il s'agissait en l'occurrence de la même version que le test subséquent - puis à un entraînement, le groupe II (n = 47) ne suivit que l'entraînement et le groupe III (n = 49) uniquement le test préalable. Le groupe IV (n = 202) passa sans préparation le test subséquent et le groupe V (n = 45) étudia la brochure sur le test à titre de préparation. Le programme d'entraînement comprenait une série d'exercices-tests, en tout 348 problèmes en 17 groupes répartis sur deux séances d'environ 3 heures chacune. Aux personnes ayant suivi l'entraînement, on dispensa en outre des instructions ciblées, des conseils spéciaux sur la manière de travailler, des aides stratégiques et des explications sur le mode de construction des problèmes, ce qui devait leur permettre d'atteindre le plus haut degré possible de "test-wisness". Résultats (Deter 1980, p. 57): les personnes n'ayant pas passé le test préliminaire mais qui, en guise de préparation, avaient ou bien suivi l'entraînement, ou bien étudié la brochure sur le test, supplantèrent par un gain de 7,9% les candidats non préparés - là non plus, pas de gain supplémentaire notable à créditer aux cours d'entraînement comparés à l'étude de la brochure.

*Influence des moyens de préparation sur les résultats obtenus au test dans l'expérience de Deter (1980, 1982)*

| Groupe | n   | Test préliminaire | Entraînement | Test subséquent | Gain* |
|--------|-----|-------------------|--------------|-----------------|-------|
| I      | 96  | oui               | oui          | oui             | 11,7% |
| II     | 47  |                   | oui          | oui             | 7,9%  |
| III    | 49  | oui               |              | oui             | 9,6%  |
| IV     | 202 |                   |              | oui             | 0,0%  |
| V      | 45  |                   | brochure     | oui             | 7,9%  |

\* Gain en % en comparaison des candidats non préparés (groupe IV)

Un groupe d'évaluation (Bartussek et al. 1984, 1986) indépendant du concepteur du test étudia de la même manière l'influence des possibilités de préparation sur les performances au TMS. En tout, 26 candidats à l'expérience (10 étudiantes,

16 étudiants) participèrent à un test préalable, à un entraînement au test et à un test subséquent tandis qu'un groupe de contrôle (n = 33, 16 étudiantes, 17 étudiants) ne passait que le test préliminaire et le test subséquent. Entre les deux tests, le groupe d'essai suivit un entraînement de 50 heures sous la conduite de deux psychologues diplômés. Résultats: le groupe d'évaluation (Bartussek et al. 1986, p. 9/11) constate qu'après l'étude de la brochure sur le test, le travail sur deux versions du TMS apporte une amélioration moyenne de quelque cinq points et le même travail sur deux versions du test, auquel s'ajoute un entraînement, améliore en moyenne la performance d'environ 15 points. L'adjonction de l'entraînement a donc apporté dix points supplémentaires, ce qui fait environ la moitié d'un écart standard. L'avantage statistiquement important réalisé par le groupe d'essai s'est manifesté dans les sous-tests "concentration au travail" (3,6 points), "figures tubulaires" (2,2 points), "retenir des figures" (2,0 points) et "compréhension de textes" (1,6 point). Pour le groupe de contrôle, ce n'est que dans le sous-test "concentration au travail" que l'on a constaté une montée statistiquement importante de la performance due à l'effet de répétition. Les résultats de cette étude ont servi à élaborer la brochure sur le test, notamment les indications au sujet de l'entraînement.

L'étude faite sur la possibilité de s'exercer à "un travail concentré et soigneux" (Fay 1989) a montré que la simple répétition de ce sous-test permet une amélioration de la performance pouvant atteindre 70%. Si de surcroît le candidat observe les conseils donnés dans la brochure à propos de ce sous-test, il parvient à doubler sa performance initiale. Des déclassements sont sans lien avec cela. Seul est désavantagé le candidat qui ne s'est pas préparé.

Dans l'étude de Klieme & Espey (1992), les indications contenues dans la brochure à propos du sous-test "retenir des faits" ont servi de base à l'entraînement appliqué. Les gains de performance pour 78 participants représentaient en moyenne la moitié d'un écart standard.

## Conclusions

En résumant le résultat des enquêtes quant à l'influence de l'exercice et de l'entraînement sur les performances réalisées dans un test standardisé, on arrive aux conclusions suivantes (cf. aussi Deter 1982, p. 68 ss):

La répétition du test aboutit - en faisant la moyenne du groupe étudié - à une augmentation plus ou moins marquée de la performance. Le conseil donné, à titre de préparation standard, de parcourir une fois le test à fond, dans des conditions proches de la réalité, est donc important. En Suisse, une version originale du test rendue publique a été traduite en français et en italien. Les trois groupes linguistiques disposent ainsi d'une base de préparation.

On a étudié, expérience à l'appui, quelle influence le matériel auxiliaire remis gratuitement aux candidats pour se préparer au test peut avoir sur les perfor-

mances. Résultat: Ceux qui ont utilisé cette documentation ont obtenu en moyenne les mêmes résultats que les candidats ayant suivi un entraînement supplémentaire.

D'une part il est donc nécessaire que les candidats étudient attentivement le matériel de préparation et qu'ils suivent les recommandations standardisées contenues dans la brochure sur le test. La preuve est donnée que ce mode de préparation assure de meilleurs résultats. Entre dans cette préparation standardisée le parcours en temps réel d'une version complète du TMS. En tout, ce travail de préparation ne dépasse pas les limites du raisonnable.

A noter d'autre part que, de toutes les études faites, on ne peut tirer **aucune indication** prouvant qu'**un entraînement supplémentaire** à la préparation standard améliorerait les performances. Il est donc faux de prétendre que la réussite du test exige un investissement démesuré en temps et en argent.

Si un test devait être introduit en Suisse, on pourrait conseiller aux candidats de s'y préparer comme le préconise la brochure sur le test. Il serait envisageable que les écoles, par exemple, organisent une "séance d'essai" qui permettrait aux candidats de passer "comme en situation réelle" une version publiée du test. Expérience faite, il n'est pas facile de travailler au test tout en veillant aux temps impartis. Cette épreuve à l'essai ne devrait pas engendrer un gros supplément de dépenses.

Partout où cela est possible, il y a lieu d'empêcher un déploiement d'offres d'entraînement à l'adresse des candidats et la commercialisation de cours supplémentaires ou autres mesures de préparation payantes.

## Bibliographie

- Allalouf, Avi (1996). The effect of coaching on the predictive validity of scholastic aptitude tests. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1996.
- Bartussek, D; Raatz, U; Schneider, B & Stapf, KH (1984). Die Evaluation des "Tests für medizinische Studiengänge". Erster Zwischenbericht. Bonn
- Bartussek, D; Raatz, U; Stapf, KH & Schneider, B (1986). Die Evaluation des "Tests für medizinische Studiengänge". Zweiter Zwischenbericht. Bonn
- Deidesheimer Kreis (1993). Feststellung der Studieneignung im Rahmen der Hochschulzulassung. Studienfeldbezogene Verfahren zur Feststellung der Studieneignung bei Hochschulzulassungsentscheidungen in Numerus-clausus- und anderen Studienfächern. Bonn: Bericht für das Bundesministerium für Bildung und Wissenschaft.
- Deter, Bernhard (1980). Übbarkeit von Leistungen im TMS. In G. Trost et al. Modellversuch "Test für medizinische Studiengänge". Vierter Arbeitsbericht: 1. Januar 1980 bis 31. August 1980. Bonn: Institut für Test- und Begabungsforschung.
- Deter, Bernhard (1982). Zum Einfluss von Übung und Training auf die Leistung im "Test für medizinische Studiengänge" (TMS). Braunschweig: Agentur Pedersen.

- Fay, Ernst (1985). Vorbereitungsmöglichkeiten auf den Test: Was gibt es? Wie wird es genutzt? Nutzt es? In G. Trost et al. Modellversuch "Tests für medizinische Studiengänge". Auswertungen zum achten und neunten Testtermin und Ergebnisse weiterer Begleituntersuchungen zum Test. Zehnter Arbeitsbericht: 1. April 1984 bis 30. September 1985. Bonn: Institut für Test- und Begabungsforschung.
- Fay, Ernst & Freitag, Gerd (1989). Über die Übbarkeit der Leistung im Test "Konzentriertes und sorgfältiges Arbeiten". In G. Trost (Hrsg.) Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 13. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Kirchenkamp, Thomas & Mispelkamp, Harald (1988). Beziehungen zwischen Leistungen im Test für medizinische Studiengänge und verschiedenen Vorbereitungsmaßnahmen, Einstellungen zum Vergabeverfahren sowie links- bzw. rechtshändiger Schreibweise. In G. Trost (Hrsg.) Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 12. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Klieme, Eckhard & Espey, Jürgen (1992). Trainingseffekte beim Merkfähigkeitstest "Fakten lernen". In G. Trost (Hrsg.) Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 16. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Mispelkamp, Harald (1987). Training für den TMS: Darstellung und Bewertung von Trainingsmaterialien, Bearbeitungshinweisen und formalen Lösungsstrategien. In G. Trost et al. Test für medizinische Studiengänge (TMS): Zehnter und elfter Termin des Übungsverfahrens. Erster Termin im besonderen Auswahlverfahren. Weitere Untersuchungen zum Test. 11. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Nauels, Heinz-Ulrich (1991). Angebote zur Vorbereitung auf den "Test für medizinische Studiengänge": Neues auf dem Bücher- und Trainingsmarkt. In G. Trost (Hrsg.) Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 15. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Trost, Günter (Hrsg.) (1977-1995). Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 1. - 19. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Zentralstelle für die Vergabe von Studienplätzen (Hrsg.) (1995). Test-Info, Herbst 94. Dortmund: ZVS.



# Qualités requises d'une procédure d'admission aux études de médecine en Suisse: lignes directrices pour la mise au point d'un diagnostic d'aptitude

U. Schallberger

## 1. Introduction et préliminaire

A supposer qu'en Suisse aussi, un *numerus clausus* aux études de médecine devienne bientôt inévitable, il faudra alors envisager la mise en place d'une procédure d'admission bien conçue. Il ressort clairement des exposés précédents que bien d'autres pays, certains depuis longtemps déjà, sont confrontés au même problème et s'emploient à le résoudre par les méthodes les plus diverses. Dans la discussion de ces méthodes on a parlé d'une série de critères de qualité auxquels elles doivent satisfaire. Ces critères de qualité sont tirés du diagnostic psychologique d'aptitudes. Nous nous proposons ici d'explorer cette perspective diagnostique et de l'appliquer à l'état de la question en Suisse. Pour anticiper la principale conclusion: dans la situation présente, le "Test d'aptitude aux études de médecine" (EMS) exposé dans le rapport de K.-D. Hänsgen représente sans conteste la solution optimale. Afin d'éviter des malentendus, il faut reconnaître que d'autres manières de voir, pouvant aboutir à une autre conclusion, sont bien évidemment possibles. La thèse ici défendue est celle-ci: si l'on veut conférer à la procédure d'admission le caractère d'un diagnostic d'aptitude, l'EMS représente, dans les données actuelles, le meilleur compromis entre le souhaitable et le réalisable.

Afin d'explicitier les réflexions qui débouchent sur cette conclusion, nous procéderons ainsi: Nous aborderons d'abord brièvement l'arrière-plan de la perspective diagnostique (point 2) puis nous esquisserons quelques conditions cadres auxquelles doit satisfaire la procédure d'admission recherchée; on verra aussi à ce chapitre pourquoi cette thèse axée sur le diagnostic d'aptitude est importante à cet égard (point 3). Les paragraphes suivants servent à expliquer cette manière de voir et ses conséquences pour le problème posé.

## 2. L'ubiquité du "diagnostic d'aptitude" et le sens d'une scientification de procédures équivalentes

Si l'on entend par diagnostic d'aptitude la tentative de constater (de "diagnostiquer") l'aptitude d'une personne à une activité **avant** qu'elle ne l'exerce, il s'agit là, naturellement, d'un phénomène ubiquitaire. C'est ainsi que, par exemple, tous les instruments de sélection dans le système d'éducation (y

compris l'attribution de notes, dès lors qu'est en jeu une promotion ou l'admission à une fonction) ont des composantes diagnostiques. Cela vaut aussi bien pour le monde du travail, où un système de "numerus clausus" a toujours existé et se manifeste à l'extrême aujourd'hui. Si le nombre de candidats excède celui des places de formation ou des emplois, on sélectionne, qu'il s'agisse d'apprentissages ou de postes de management à l'échelon supérieur. Les instruments utilisés à cet effet sont aussi divers que variés, allant de l'examen du dossier du candidat en passant par l'entretien d'embauche, jusqu'à des expertises de plusieurs jours, sans parler de la méthode plus coûteuse encore de l'"Assessment Center". Tout cela pour faire le bon choix en fonction de l'aptitude des candidats.

Pour légitimer une méthode concrète de ce genre, il y a en principe deux manières de procéder. Nous avons d'un côté des argumentations au nom du principe dit "per fiat"; exprimé dans le langage courant: "Il est évident pour moi que cette méthode livre l'information essentielle; elle le fera donc!" Dans l'autre perspective, une méthode n'est valable et défendable que moyennant la preuve empirique qu'elle prend effectivement en considération l'aptitude. Ce second point de vue constitue l'attitude fondamentale dans un diagnostic psychologique d'aptitude scientifiquement fondé. Cette attitude peut se réclamer d'une foule d'exemples dans lesquels des méthodes "per fiat" de vérification empirique par des études de longue durée se sont révélées douteuses, carrément inutilisables, voire contraires au but visé. La recherche des causes de ce constat totalement surprenant a donné lieu à une multitude d'incursions dans les difficultés de cette entreprise risquée qu'est le diagnostic d'aptitude. Il en est sorti une discipline scientifique en soi, le diagnostic psychologique d'aptitude, qui produit aussi les outils théoriques et méthodiques nécessaires à la mise au point de procédures de sélection valables. Tout cela du reste n'a pas qu'un intérêt purement académique: le rapport coût/utilité d'une procédure de sélection ne peut être effectivement optimisé que sur ce fond scientifique<sup>1</sup>.

La "quintessence", si l'on peut dire, de cette perspective diagnostique est une liste de "critères de qualité" pour la procédure de diagnostic d'aptitude, que G. Trost a présentés en partie dans son rapport. La majorité de ces critères ne sont d'ailleurs qu'imparfaitement réalisables. Mais il est admis que l'on doit vérifier en tous cas dans quelle mesure ces critères sont remplis, de manière à ce que les points négatifs et positifs d'une procédure soient ouvertement déclarés et donnent lieu à une discussion objective. Nous revenons aux chapitres 4ss sur la justification de ces critères et sur leur pertinence.

---

<sup>1</sup>Les avantages d'une procédure de sélection ne sont abordés ci-après qu'en marge. On trouve une brève introduction à ce sujet dans Schallberger, U. (1996). Nutzen, Fairness, Validität und Akzeptanz von Selektionsverfahren. Dans K.-D. Hänsgen et al., Eignungsdiagnostik und Medizinstudium (p. 38-42). Rapport 2 du Centre pour le développement de tests et le diagnostic de l'Université de Fribourg.

### 3. Conditions et exigences cadres

Il faut d'abord se remémorer encore une fois les conditions et les exigences cadres auxquelles doit satisfaire, dans notre cas, la procédure d'admission étudiée. Car il est vain d'imaginer des procédures qui seraient de toute façon irréalisables dans le cadre donné. Nous présentons et mettons en discussion ci-après, sous forme de thèses, les cinq conditions cadres les plus importantes:

1. **Une solution nationale serait indiquée:** Il existe certes de solides arguments en faveur de l'autonomie des universités, également dans la politique de l'admission; entre autres le fait que l'autonomie dans la sélection implique aussi une plus grande responsabilité institutionnelle envers la formation des étudiants (que l'université a elle-même choisis). Mais aujourd'hui, alors que des permutations d'étudiants entre universités sont nécessaires avant et pendant les études de médecine, une solution décentralisée se heurterait à de grandes difficultés pratiques.
2. **Il est nécessaire de traiter sur pied d'égalité tous les candidats et candidates (indépendamment du canton d'origine, de l'école, de la langue, du sexe):** Ce deuxième point se passe de justification. Dans notre système de formation il n'est guère imaginable que par exemple - comme c'est le cas aux Etats-Unis - la sélectivité (de quelque manière qu'elle soit déterminée) de l'école dont sort le candidat soit érigée en critère d'admission. L'idée d'égalité de traitement fait obstacle aussi à l'utilisation des notes de maturité ou de tests de connaissances. Le système de notation varie beaucoup selon les écoles et les régions; elle varierait sans doute plus encore si les notes de maturité prenaient de l'importance dans la sélection. Quant à l'introduction de tests de connaissances, elle ferait assez vite éclore une "industrie de l'entraînement" dont ne profiteraient que les candidats disposant des moyens financiers nécessaires.
3. **Les ressources en temps et en argent pour la mise au point d'une procédure d'admission sont extrêmement limitées:** A observer les exemples étrangers de procédures d'admission à l'université, il apparaît nettement que la mise sur pied d'une telle procédure en tenant compte des problèmes particuliers serait extrêmement laborieuse. Rappelons par exemple que l'actuelle version du MCAT américain est le fruit de 8 ans de préparatifs et de recherche (cf. rapport de J. L. Hackett). On peut en dire autant d'autres modèles. Manifestement, la Suisse n'a aujourd'hui ni le temps ni les moyens d'élaborer de toutes pièces sa propre solution.
4. **Les ressources disponibles pour l'organisation de la procédure d'admission sont limitées:** Ce fait restreint aussi notre marge de manoeuvre. Vu la précarité des ressources, il ne serait pas réaliste, par exemple, de mettre sur pied pour tous les intéressés aux études des entretiens de sélection individuels. Tout au plus pourrait-on envisager un système selon lequel un

grand nombre de candidats passeraient simultanément un test ou un examen, les interviews individuelles étant réservées à un petit nombre de cas spéciaux ou de cas limites.

5. **Le nouveau système d'admission doit prendre en considération l'aptitude du candidat mieux que ce n'est le cas dans le système actuel de l'autosélection:** Actuellement, l'accès aux études de médecine est fondé sur **l'autosélection**, sous réserve d'avoir obtenu la maturité. La personne qui s'intéresse aux études de médecine et s'estime capable de les suivre est admise. La **sélection par des tiers** n'a lieu qu'aux examens propédeutiques. Autant qu'on sache, tous les milieux intéressés voudraient qu'en plus de l'autosélection la nouvelle procédure d'admission prenne en compte plus objectivement l'aptitude aux études. Il y aurait bien une solution parfaite au problème de l'admission: la **loterie**, soit la sélection par le hasard. Elle satisferait presque entièrement aux quatre exigences précitées (réalisable à l'échelon national, égalité de traitement, modicité des frais de développement et d'exécution). Comme l'a montré le tableau de G. Trost, cette méthode ne joue qu'un rôle insignifiant sur le plan international, et c'est compréhensible: comment pourrait-on laisser le sort décider de l'admission aux études? L'idée est dès lors la suivante: puisque les places d'études sont déjà limitées, il semble juste et raisonnable - sous l'angle des pouvoirs publics, des universités et des étudiants - de donner systématiquement la préférence aux candidates et candidats dont l'aptitude est la mieux établie (le terme d'aptitude pouvant, bien sûr, recouvrir diverses qualités). Le principe de l'égalité est ici sauvegardé, bien que modifié au sens d'un traitement égal pour les candidats possédant la même aptitude. On parle souvent d'équité pour désigner ce principe ainsi modifié de l'égalité de traitement.

S'il convient de retenir l'aptitude en tant que critère d'admission, la procédure d'admission devient d'office une procédure de diagnostic d'aptitude au sens décrit plus haut - avec toutes les difficultés liées à une telle méthode.

#### 4. De la structure fondamentale des énoncés du diagnostic d'aptitude

Pour se faire une idée des sources de difficulté du diagnostic d'aptitude et des possibilités de les maîtriser, il faut aborder en premier lieu la structure de base d'un énoncé diagnostique.

Un énoncé du diagnostic d'aptitude a, par nature, le caractère de **pronostic**: sur la base d'informations momentanément disponibles, appelées **prédicats**, un avis sera prononcé sur l'ampleur du succès prévu, appelé **critère**. Un tel énoncé est dans tous les cas - exactement comme les pronostics du météorologue ou du spécialiste en conjoncture - entaché d'incertitudes: c'est un énoncé de

probabilité. Les facteurs d'incertitude sont de nature extrêmement diverse, comme le montre l'illustration:

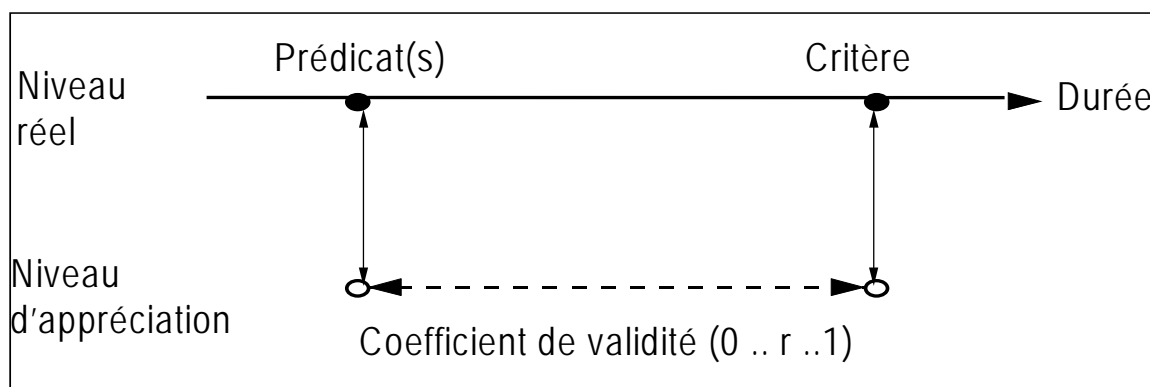


Illustration 1 : Structure d'un énoncé de diagnostic d'aptitude

Dans ce tableau, la ligne horizontale représente l'axe de durée où se localisent les prédicats et le critère. Sur la verticale, deux niveaux: le "niveau de réalité" et le "niveau d'appréciation". Cela peut s'expliquer plus simplement par un exemple:

Supposons que le critère à pronostiquer soit "bon médecin" et le prédicat à appliquer "empathie", cela dans l'idée qu'un bon médecin doit être capable d'un minimum d'empathie envers le patient. Dans le schéma, cette réflexion concerne le "niveau de réalité". Elle renferme des énoncés, ou mieux des hypothèses sur la réalité. Pour en tirer profit dans une procédure de diagnostic d'aptitude, ces énoncés doivent encore être transformés au niveau dit "d'appréciation" prévu dans le schéma. A cet effet il faut choisir une procédure qui permette de porter une appréciation sur la capacité empathique du candidat; seule une telle appréciation pourra effectivement être utilisée comme prédicat. S'il faut en outre vérifier la qualité du pronostic (et cela **doit** toujours se faire dans la perspective d'un diagnostic d'aptitude, cf. paragraphe 2 ci-dessus) il faudra une procédure d'appréciation permettant de distinguer entre "bons médecins" et "moins bons médecins" ("critère").

Si dans ce sens le ou les prédicats et le critère passent au "niveau d'appréciation" il sera possible après un certain temps de faire le rapport entre le prédicat et le critère. Le résultat mesure la qualité de la prévision, le degré d'incertitude et l'étendue des fausses appréciations dus à la procédure choisie. Le coefficient y relatif est appelé **validité pronostique**; on l'a déjà vu dans plusieurs des précédents rapports. Ce coefficient varie en général entre 0 et 1. "1" indique une parfaite prédiction - qui en réalité ne peut jamais être atteinte. "0" signifie que la "qualité" du pronostic correspond exactement à celle de la loterie. Le coefficient serait négatif si la procédure d'admission donnait systématiquement la préférence aux candidats les moins aptes, ce qui est certes totalement indésirable mais se produit pourtant régulièrement dans la réalité.

## 5. Déterminants de la validité pronostique d'un énoncé de diagnostic d'aptitude

Le schéma de l'illustration 1 permet aussi de rassembler les déterminants de la validité pronostique d'une prévision de diagnostic d'aptitude. Ces déterminants sont indépendants de l'exemple cité; ils valent pour n'importe quels prédicats et critères auxquels on fait appel pour résoudre un quelconque problème de diagnostic d'aptitude.

### 1. Rôle effectif du prédicat pour le critère au "niveau de la réalité"

A ce stade, certaines questions se posent: Le rapport requis entre prédicat et critère est-il valable? N'y a-t-il pas d'autres facteurs qui entrent en ligne de compte et qui jouent même un grand rôle? Il devrait être évident d'emblée que le nombre et l'influence de ces autres facteurs ont une influence négative sur la validité pronostique d'un prédicat.

2. **Espace de temps entre prédicat et critère:** Etant donné qu'une personne est susceptible d'évoluer dans l'une ou l'autre des directions, la validité pronostique est par nature d'autant plus faible que l'espace de temps est long entre le prédicat et le critère. Cette observation a une importance particulière dans la sélection pour les études de médecine car il s'agit de jeunes personnes, engagées dans une formation qui - suivant les circonstances - pourrait elle-même connaître toute une série d'évolutions.

### 3. Objectivité, fiabilité et validité du prédicat

Le point central est ici la validité de l'appréciation sur le prédicat, qu'il faut distinguer de la validité pronostique. Il y a cependant un rapport entre les deux. Plus faible est la validité de l'appréciation, plus faible est aussi la validité pronostique de la procédure. Pour reprendre l'exemple ci-dessus: Si la capacité d'empathie d'une personne n'est pas valablement perçue, il est clair que l'appréciation y relative ne peut pas non plus avoir une validité pronostique substantielle. Les deux autres critères - objectivité et fiabilité - (que G. Trost a définis dans son rapport) acquièrent de l'importance, tout insuffisants qu'ils soient, parce qu'ils désignent les conditions nécessaires à la validité de l'appréciation. Par exemple, une appréciation qui dépend plus de son auteur que de la personne jugée (= manque d'objectivité) ne peut a priori être valable. Ces observations permettent au demeurant de comparer entre elles diverses techniques de procédure, comme l'a fait G. Trost dans son rapport. Ainsi, il ressort de nombreuses études de contrôle que des appréciations données à l'issue d'entretiens menés à la manière d'interviews, n'ont que peu d'objectivité et de fiabilité, et n'acquièrent de ce fait qu'une faible validité pronostique. Seules des mesures standardisées, assez strictes (questions fixées d'avance, schémas d'appréciation pré-déterminés, instruction pour la conduite des interviews) permettent de déduire des interviews des prédicats potentiellement utilisables.

Diverses catégories de prédicats peuvent être évalués de manière analogue. Ainsi, il s'est révélé beaucoup plus difficile d'acquérir des appréciations objectives, fiables et valables sur des qualités morales que ce n'est le cas pour des aptitudes. C'est sur des caractéristiques relativement proches du comportement que les perspectives sont le plus favorables.

#### 4. Objectivité, fiabilité et validité du critère

Ce qui vient d'être dit pour le prédicat vaut *mutatis mutandis* pour le critère. Du point de vue du diagnostic d'aptitude il n'y a guère de sens à vouloir prédire un critère qui ne peut être saisi de manière objective, fiable et valable du fait que les pronostics sont pratiquement invérifiables. Dans ces conditions, une justification de la procédure est une pure question de foi.

En se remémorant tous ces déterminants de la validité pronostique d'une procédure de sélection, on comprendra peut-être pourquoi un tel instrument est si long et difficile à développer comme on l'a vu plus haut. Chacun des aspects évoqués demande à être élucidé de manière empirique avant qu'une procédure ne soit légitimement reconnue rationnelle.

Au terme de ces réflexions, mentionnons encore une manière de procéder qui, il est vrai, n'élimine pas toutes les difficultés des points 1, 3 et 4, mais tout au moins les atténue quelque peu. Il s'agit de la différence entre procédures **orientées sur l'aptitude** et procédures **orientées sur la simulation**. L'exemple de la sélection d'un pilote nous montre en quoi consiste cette différence. Serait orientée sur l'aptitude, une procédure qui utilise comme prédicats des qualités dont on admet qu'elles sont nécessaires à un "bon pilote" (par exemple la capacité de représentation spatiale, la concentration, la résistance au stress, etc.). L'appréciation du degré de ces qualités se fait au moyen d'instruments standardisés, non spécifiques aux pilotes. Serait en revanche orientée sur la simulation, une procédure fondée le plus fidèlement possible sur les exigences posées au pilote dans le quotidien, par exemple en rapport avec les capacités de concentration ou la résistance au stress. La procédure utilise alors comme prédicats des performances qui simulent aussi bien que possible ces exigences spécifiques. "Aussi bien que possible" ne se rapporte pas à l'examen mais à la parenté des structures psychologiques nécessaires lors de la situation simulée et de la situation réelle. Si elles sont construites selon les règles de l'art, les procédures orientées sur la simulation ont en général plus de validité pronostique que celles orientées sur l'aptitude, cela avant tout pour deux raisons: d'abord parce que le rapport entre prédicat et critère est a priori moins spéculatif. Ensuite parce que cette procédure (orientée sur la simulation) permet de formuler le prédicat et le critère sous une forme plus proche du comportement. Exprimé dans les termes de l'illustration 1: La distance entre le "niveau

d'appréciation" et le "niveau de la réalité" ainsi que les problèmes qui y sont liés se trouvent minimisés.

## 6. Conclusions

Au départ de notre réflexion, nous avons considéré qu'un *numerus clausus* aux études de médecine pourrait bientôt être inévitable en Suisse aussi, ce qui nécessitera la mise en place d'une procédure d'admission bien conçue. Etant donné que cette procédure devrait aussi avoir une fonction de diagnostic d'aptitude, nous avons abordé quelques problèmes essentiels liés à un tel objectif. Si l'on confronte maintenant nos observations avec la situation actuelle en Suisse, on s'aperçoit très vite qu'au point de vue du diagnostic d'aptitude, il n'existe pour le moment aucune autre solution utilisable que le "Test d'aptitude aux études de médecine" (EMS) présenté dans le rapport précédent. Les principaux arguments qui aboutissent à cette conclusion sont résumés ci-après:

- La mise au point d'une procédure de diagnostic d'aptitude qui tienne compte des problèmes évoqués exigerait des ressources en temps et en argent dont on ne dispose pas en Suisse à l'heure actuelle; d'où la nécessité de se référer à des travaux réalisés à l'étranger. Pour s'assurer d'un bon fonctionnement, il convient de choisir cet instrument dans un pays dont le système universitaire a le plus d'analogie possible avec le système suisse. L'EMS, qui a été mis au point et contrôlé en Allemagne durant de nombreuses années, est la seule procédure qui réponde à ce critère.
- Le test EMS satisfait au mieux aux exigences cadres que nous avons esquissées. Il est rapidement disponible, réalisable à l'échelle nationale et en conformité avec les principes d'économie. En outre, des études empiriques attestent qu'il garantit la plus large équité.
- Le test EMS utilise un critère de résultat - le succès dans les études - qui peut être qualifié de réaliste pour une procédure de diagnostic d'aptitude. L'espace de temps entre prédicat et critère reste ainsi dans un cadre raisonnable. D'une perspective supérieure, on aurait pu préférer un critère lié à la future activité professionnelle de médecin. Mais c'eût été au prix d'un important travail supplémentaire et surtout, le problème de prédictivité en serait devenu beaucoup plus complexe, aboutissant à une validité pronostique à peine supérieure à celle d'un tirage au sort.
- Le test EMS est une procédure orientée sur la simulation; sa construction a nécessité de grands travaux préliminaires. Il vise à simuler des exigences cognitives jugées importantes et auxquelles les étudiants sont confrontés au cours de leurs études. Cela rend le détour, d'ailleurs semé d'embûches, par des hypothèses sur les qualités indispensables au candidat, superflu.



- La validité pronostique de l'EMS est en conséquence à la limite supérieure de ce qui peut être atteint par un mode de sélection de ce genre et à conditions sélectives comparables. En Allemagne, les coefficients de validité se situent à environ .5, valeur que l'on peut qualifier d'élevée selon les normes internationales usuelles et qui n'est jamais sensiblement dépassée dans des conditions égales et selon les mêmes méthodes de calcul.

L'EMS souffre d'un point faible qu'il ne faut pas passer sous silence: des enquêtes de contrôle en Suisse, notamment sur la validité pronostique du test, font encore largement défaut. La référence aux résultats allemands ne devrait pas devenir un argument "per fiat". Si l'EMS est effectivement introduit, il sera indispensable de l'accompagner d'études empiriques bien conçues et d'avoir à disposition les fonds nécessaires à cet effet (pour des cas analogues dans le monde du travail, les entreprises avisées considèrent de telles vérifications comme allant de soi). C'est à cette condition que l'on pourra juger objectivement si et dans quelle mesure la procédure choisie est préférable à un tirage au sort. Au niveau du diagnostic d'aptitude, seul un avantage notoire justifiera le déploiement de moyens utiles à l'Université comme aux candidats aux études concernés par le *numerus clausus*.

## Adresses

Dr. John L. HACKETT

---

MCAT Section

Association of American Medical Colleges; 2450 N Street, NW;

**Washington, DC 20037**

U. S. A.

PD Dr. Klaus-Dieter HÄNSGEN; lic.phil. Rainer HOFER

---

Zentrum für Testentwicklung und Diagnostik

am Psychologischen Institut der Universität Freiburg

Rte d'Englisberg 9

**CH-1763 Granges-Paccot**

Prof. Dr. Widar HENRIKSSON; Prof. Dr. Ingemar WEDMAN

---

Department of Educational Measurement

University of Umeå

**S-Umeå 90187**

Prof. Dr. Piet J. JANSSEN

---

Centrum voor Schoolpsychologie

Departement Psychologie

Katholieke Universiteit Leuven

Tiensestraat 102

**B-3000 Leuven**

Prof. Dr. Urs SCHALLBERGER

---

Psychologisches Institut

Universität Zürich

Schönberggasse 2

**CH-8001 Zürich**

Dr. Günter TROST

---

Institut für Test- und Begabungsforschung

Koblenzerstrasse 77

**D-53177 Bonn**