



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

# **Le test d'aptitudes pour les études de médecine en Suisse**

*Epreuve d'essai*

Rainer Hofer, Daniel Ruefli & Klaus-Dieter Hänsgen

Rapport 1 (1995)

# Table des matières

<b>RÉSUMÉ .....</b>	<b>3</b>
<b>EXPOSÉ DU PROBLÈME.....</b>	<b>5</b>
<b>MÉTHODE.....</b>	<b>6</b>
Echantillon.....	6
Instruments et déroulement du test .....	6
<i>Préparation.....</i>	7
<i>Questionnaire 1.....</i>	7
<i>Le Test.....</i>	8
<i>Questionnaire 2.....</i>	8
<i>Groupe témoin .....</i>	9
<i>Déroulement du Test.....</i>	9
<b>RÉSULTATS.....</b>	<b>11</b>
Description de l'échantillon .....	11
<i>La difficulté du test.....</i>	12
<i>Précision des mesures.....</i>	13
<i>Relations entre les sous-tests et les notes scolaires.....</i>	13
<i>Relation entre le sexe et les résultats au test .....</i>	14
<i>Relation entre la langue et le résultat au test .....</i>	18
<i>Résultat au test et autres critères.....</i>	20
<b>DISCUSSION .....</b>	<b>27</b>
<i>Le test satisfait aux critères de qualité .....</i>	27
<i>Le test permet une sélection optimale.....</i>	27
<i>Le test ne mesure pas la même chose que les notes scolaires .....</i>	28
<i>Le test ne défavorise pas le femmes .....</i>	29
<i>L'égalité des chances vaut aussi pour les groupes linguistiques .....</i>	30
<i>Conclusion .....</i>	30
<b>LITTÉRATURE .....</b>	<b>31</b>

---

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  
Tel. 037/ 29 79 86 - Fax 037/ 29 97 63 - e-mail: ZTD@UNIFR.CH  
Internet: <http://www.unifr.ch/Main/ztd.html>

## Résumé

L'introduction d'un numerus clausus en Suisse pour l'admission aux études de médecine relève d'une décision politique, qui doit être prise en tenant compte de différents facteurs. A dire vrai, on ne pourra faire admettre un numerus clausus que si l'on dispose d'un critère reconnu, sûr et si possible équitable pour régler cette admission. En ce qui concerne les autres points de la discussion sur ce thème, il convient d'observer avec une attention redoublée

- qu'il appartient aux milieux politiques de se prononcer sur la nécessité d'une limitation de l'accès aux études par le biais d'un numerus clausus et de définir la capacité d'accueil des hautes écoles et universités;
- que, pour fixer **la condition sous-jacente à ce numerus clausus**, il faut trouver un critère qui soit relativement équitable ou le plus équitable possible.

Aujourd'hui comme hier, on vise à coordonner dans l'ensemble de la Suisse l'admission aux études de médecine. Jusqu'à maintenant, aucune expérience n'avait été tentée pour savoir s'il existe en Suisse un test d'aptitudes qui puisse être appliqué à l'instar de ce qui se fait en Allemagne. Chez nous, par exemple, le test doit pouvoir être conduit dans les trois langues nationales, ce qui n'est pas le cas en Allemagne. La nécessité absolue de garantir l'égalité des chances a toujours été au centre des discussions sur le test et un certain nombre de craintes ont été émises à ce sujet.

L'organisation d'un test en Suisse devait dès lors servir plus particulièrement à élucider les questions suivantes:

- **La sélectivité** en fonction de la performance est-elle vraiment un critère d'admission adéquat ? Ou bien faut-il admettre, d'une manière générale, que le test est trop facile ou trop difficile pour qu'il soit possible d'opérer une sélection sur la base des résultats obtenus?
- Est-ce que les **formes linguistiques** du test sont équivalentes et garantissent la même fidélité et la même équité pour tous les groupes linguistiques?
- Les performances réalisées lors des tests mettent-elles en évidence des différences entre les **sexes** ou les **différents types de maturité**, telles qu'elles excluent la possibilité d'une égalité des chances ?

Afin de vérifier les critères de qualité psychologiques et l'objectivité du test, l'on a organisé au collège de Sainte-Croix à Fribourg (Suisse) une épreuve réunissant un échantillon de 54 lycéennes et lycéens de langue allemande et 126 de langue française, puis l'on a comparé leurs résultats avec ceux obtenus par le groupe témoin allemand lors de la passation du test en 1992.

Il faut observer ici que, lors de cette épreuve d'essai, les participantes et participants n'étaient pas véritablement en situation de candidature au sens où on l'entend habituellement, la seule "motivation" qui pouvait leur être offerte consistant à se voir communiquer, en retour, leur résultat individuel au test.

Nonobstant ce manque de véritable motivation chez les lycéens de l'échantillon fribourgeois, le test a atteint, du point de vue de la fiabilité **des critères de qualité** comparables à ceux obtenus en Allemagne. En moyenne, la moitié des exercices cotés (48%) ont été résolus correctement et la distribution des résultats se situe dans un champ optimal pour la **différenciation des performances**.

Les **groupes linguistiques** ont réalisé tous deux de **bonnes valeurs de fiabilité**, ce qui parle en faveur d'une grande équivalence des deux versions rédactionnelles du test.

Une question importante demeure celle du comportement des sexes par rapport aux performances réalisées au cours du test. L'hypothèse émise sur la base des résultats obtenus en Allemagne selon laquelle **les femmes ont tendance à réaliser de moins bonnes performances** que les hommes **n'a pas été confirmée en Suisse**. Dans le groupe de langue allemande, les femmes ont même obtenu de meilleurs résultats que les hommes. Pour ce qui concerne le groupe francophone, on n'a enregistré aucune différence significative au niveau des résultats des deux sexes.

Si l'on examine les résultats au test pour juger du succès futur des examens préalables aux études de médecine, l'on constate que, dans le groupe francophone, les valeurs moyennes du test obtenues par les élèves qui ont opté pour la maturité de type D sont significativement plus faibles que celles réalisées par ceux qui ont choisi les types A, B et C. Ce constat semble confirmer les résultats des examens de médecine à Berne, où 44 % à 50 % des étudiantes et étudiants au bénéfice d'une maturité de type A à C échouent lors de leur premier essai au premier examen propédeutique, contre 77 % pour les détenteurs de type D (Hofer 1992).

Dans l'ensemble, les résultats de l'épreuve ont montré à l'évidence **que le test pourrait être utilisé également en Suisse en tant que critère de sélection équitable et sûr**. Au surplus, ils n'ont pas confirmé les appr-

hensions que l'on pouvait avoir au sujet de la différence des performances entre les sexes, les groupes linguistiques ou même par rapport à l'Allemagne, pays où le test a été développé.

## Exposé du problème

L'introduction d'un *numerus clausus* exige un critère de sélection objectif, reconnu, qui règle l'admission aux études supérieures. Ce critère doit être juste (équitable) et garantir l'égalité de traitement de tous les groupes linguistiques, des deux sexes, etc. Pour limiter l'accès aux études de médecine en Suisse, on prévoit d'instaurer un test d'aptitudes qui servira de critère de sélection. Ce test d'aptitudes que l'on peut décrire le plus justement comme étant une "épreuve probatoire" comportant des exercices typiques pour les études de médecine a été mis au point en Allemagne (Trost et al. 1977 à 1994) et adapté aux conditions suisses. Ses résultats permettent de formuler un pronostic sur l'aptitude aux études des différentes personnes testées. En tant qu'instrument de prévision quant au succès des études, ce test compte au nombre des meilleures procédures par comparaison internationale.

Dans les discussions qui ont eu lieu jusqu'ici en Suisse à propos de l'introduction d'un test, on se référait aux résultats obtenus en Allemagne, parce que c'est dans ce pays que le test a été introduit en 1986 et qu'il a prouvé depuis lors sa fiabilité dans un grand nombre d'études scientifiques qui lui ont été consacrées. Mais les résultats obtenus en Allemagne ont conduit certains à affirmer que le test n'était pas suffisamment spécifique pour des études de médecine ou qu'il discriminait les femmes et les personnes provenant des "couches sociales inférieures" (Hänsgen et al. 1995).

L'objectif de la présente étude est d'examiner si les allégations de ce genre sont fondées ou si l'on a "démoli" le test en voulant s'en prendre au *numerus clausus*. Etant donné que, non seulement le test allemand a été adapté sur le plan rédactionnel pour la Suisse, mais qu'il a aussi été traduit dans les langues nationales française et italienne de ce pays, on ne devrait pas s'attendre à une similitude totale entre les résultats.

L'épreuve-test qui s'est déroulée à Fribourg a permis d'étudier les questions suivantes:

- Est-il possible d'organiser le test d'aptitudes en Suisse dans les mêmes conditions qu'en Allemagne ?
- Quelle est l'importance des paramètres qualitatifs du test d'aptitudes au cours de l'épreuve ? Présentent-ils des différences par rapport au groupe témoin allemand ?

- Note-t-on des différences selon le sexe au niveau du test, des sous-tests et des items ?
- Existe-t-il des différences d'ordre culturel ou linguistique au niveau du test, des sous-tests et des items du fait de la traduction du test en français?
- Quelle est l'influence sur les résultats au test de facteurs tels que l'âge, le type de maturité, les centres d'intérêts à l'école, les notes ou la motivation?

## **Méthode**

### ***Echantillon***

L'épreuve s'est déroulée au collège Sainte-Croix à Fribourg. Elle s'adressait à des lycéennes et lycéens qui se trouvaient à une année des examens de maturité.

Contrairement au groupe témoin allemand, ces personnes n'avaient pas la même motivation qu'un candidat. De plus, il fallait autant que possible éviter que le déroulement de l'épreuve ne perturbe le cours normal de l'enseignement; c'est pourquoi elle fut programmée à la fin de l'année scolaire, soit après la période des examens. Deux sources de motivation intrinsèque: le bénéfice de l'expérience faite et la possibilité de connaître les performances personnelles réalisées.

### ***Instruments et déroulement du test***

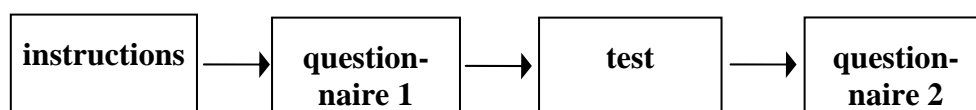
Les exercices proposés ont été empruntés à un test original qui a été publié en version allemande (Institut für Test- und Begabungsforschung 1990) sous le titre Test für Medizinische Studiengänge (TMS). Ce texte a été traduit en français (et en italien) en vue d'une publication devant servir, le cas échéant, de matériel de préparation à un test réel.

Ses résultats ont un rapport évident et prouvé avec l'aptitude aux études, puisque cette dernière a servi de critère pour la mise au point du test. Les conditions du test se rapprochent de celles des études, dans le sens d'une "épreuve probatoire". En effet, pour définir les exercices soumis aux candidats, des chargés de cours de médecine et des maîtres de lycée se sont réunis autour d'une même table, afin de juger de la représentativité de ces exercices par rapport aux exigences des études. Le test n'est pas un examen des connaissances; il entend vérifier seulement la capacité de raisonnement

et l'aptitude à résoudre des problèmes. Il n'exige pas des connaissances spéciales de la médecine, l'exercice fournissant les données (de faits) nécessaires à la résolution du problème. Il s'agit donc d'exigences spécifiques à l'aptitude à ces études.

L'application du test est équitable dans la mesure où des conditions standardisées quant à son déroulement et quant à l'analyse des résultats garantissent l'égalité des chances. L'"arbitraire" d'un examinateur ne porte pas à conséquence. Aucun critère subjectif n'entre dans l'appréciation des résultats, ce que l'on ne peut pas ou difficilement éviter par ex. dans les situations comportant une interrogation en tête à tête.

Le déroulement de l'épreuve est présenté dans le design suivant:



## PRÉPARATION

Pour se préparer au test, il importe de se familiariser avec les types d'exercices, les instructions et le déroulement de l'épreuve. C'est pourquoi, une semaine auparavant, les lycéennes et les lycéens ont été informés du déroulement du test. Les différents types d'exercices ont été discutés. On a également diffusé une brochure d'information (Centre pour le développement de tests et le diagnostic), afin de permettre aux personnes concernées de se préparer au test à domicile. Cette brochure comprend la matière qui serait utilisée pour l'application d'un test réel et qui a effectivement servi à l'épreuve de 1995.

Le manque de possibilité d'entraînement à l'épreuve-test se justifie par le fait que le test ne vérifie pas les connaissances de faits, mais porte sur la résolution de problèmes à partir des seules données figurant dans les exercices (Hänsgen et al. 1995). La diffusion préalable des conditions de l'épreuve a pour but d'éviter qu'un temps précieux ne soit consacré, le jour du test, à l'étude des instructions et de permettre ainsi aux élèves de commencer aussitôt à traiter les exercices.

## QUESTIONNAIRE 1

Lors de l'instruction préalable, on a remis aux personnes appelées à participer à l'épreuve un questionnaire destiné à obtenir divers renseignements sur leurs conditions sociales et leur éducation. Les lycéennes et les lycéens étaient libres de répondre aux questions concernant leur âge, leurs centres

d'intérêt à l'école, la moyenne des notes obtenues, le type de maturité choisi, la branche d'étude choisie ou envisagée et la formation des parents. En revanche, ils étaient invités à indiquer leur sexe et à remettre le questionnaire le jour du test.

## LE TEST

Par rapport au test original appliqué dans une situation réelle (Trost 1977 à 1994), qui se déroule en Allemagne à la date fixée entre 08.45 et 16.00 h., les lycéennes et lycéens n'ont disposé pour cette première épreuve que d'une demi-journée. C'est pourquoi ils n'ont eu à traiter que 7 des 9 sous-tests qui avaient été préparés. Le plan suivant fut établi pour le déroulement de l'épreuve, étant précisé que les points 1 et 4, resp. 2 et 5 comptaient chacun pour un sous-test.

	temps de prépa- ration en minutes	nombre d'items
1. Phase d'apprentissage du sous-test "Mémoire de figures"	4	(20)
2. Phase d'apprentissage du sous-test "Mémoire de faits"	6	(20)
3. Sous-test "Compréhension de textes"	60	24
4. Phase de reproduction du sous-test "Mémoire de figures"	5	20
5. Phase de reproduction du sous-test "Mémoire de faits"	7	20
6. Sous-test "Reconnaissance de fragments de figures"	22	24
7. Sous-test "Figures tubulaires"	15	24
8. Sous-test "Problèmes quantitatifs et formels"	60	24
9. Sous-test "Travail avec soin et concentration"	8	1200 signes = 20
<b>total expérimentation</b>	<b>187</b>	<b>156</b>

Le temps net disponible pour le test a donc été de 3 heures et 7 minutes pour résoudre 156 exercices. Afin de garantir le plus d'objectivité possible dans le déroulement de l'épreuve, les responsables du test, ont donné lecture, en allemand et en français, de tous les textes d'instructions tirés d'un manuel établi à l'attention des responsables, rédigé dans ces deux langues.

## QUESTIONNAIRE 2

Après l'exécution du test, les lycéennes et les lycéens ont été invités à répondre aux questions ci-après. Ils avaient à choisir entre les quatre modèles de réponse suivants: "inexact", "assez peu exact", "exact pour l'essentiel" et "absolument exact".

1. Le test m'a paru facile.



2. Je n'ai pas pu me concentrer pendant toute l'épreuve.
3. Je me suis senti dépassé(e) par le test.
4. J'ai trouvé les exercices pénibles.
5. J'ai eu peur de participer au test.
6. Je participerais sans appréhension à un tel test dans des conditions réelles.

Après avoir répondu à ces questions, les lycéennes et lycéens ont dû encore indiquer le pourcentage des exercices qu'ils estimaient avoir traités correctement et préciser si leurs performances avaient été meilleures ou moins bonnes que leurs notes scolaires. Les personnes testées avaient également la possibilité d'ajouter sur le questionnaire 2 des observations personnelles à propos du déroulement du test. Elles devaient également indiquer leur adresse si elles souhaitaient recevoir un avis écrit sur leur performance personnelle à titre d'information.

## **GROUPE TÉMOIN**

On a choisi comme groupe témoin les personnes qui, en 1992, ont effectué le test d'accès aux études de médecine en Allemagne. Par rapport à ce groupe témoin, l'épreuve de Fribourg ne comportait pas les mêmes exercices, mais des problèmes analogues de même type. Les nombreuses enquêtes qui ont été faites jusqu'ici sur les classes d'âge ayant participé au test en Allemagne (Trost 1977-1994) montrent qu'il est possible de comparer directement entre eux les résultats des différentes classes d'âge. Comme les exercices de la passation d'essai suisse correspondent à une version originale du test allemand, il est donc possible de comparer les résultats de l'épreuve fribourgeoise à ceux du test original allemand.

## **DÉROULEMENT DU TEST**

Les tests se sont déroulés dans des locaux séparés pour chacun des deux groupes linguistiques. La température de saison, très élevée, n'a pas facilité les conditions ambiantes. Comme les élèves étaient relativement nombreux dans chaque local, il est probable que la chaleur a quelque peu entamé la concentration au fur et à mesure du déroulement de l'épreuve. L'encadrement du groupe francophone était composé d'un responsable du test et de trois auxiliaires; le groupe germanophone, d'un responsable et de deux auxiliaires. Dans l'ensemble, on s'en est tenu à la procédure décrite dans le manuel à l'usage des responsables du test.

L'épreuve a eu lieu le 28 juin 1995, de 08.10 à 12.00 h. Un échantillon de 54 lycéennes et lycéens de langue allemande et de 126 lycéennes et lycéens

de langue française entama le test. Dans le groupe germanophone, 12 personnes ont interrompu prématurément l'épreuve. En raison des différences de motivation attendues, les responsables ont été contraints d'exclure le plus rapidement possible les participants non motivés, dès les premières manifestations de dépit. Il fallait, en effet, tenir compte des effets possibles de transfert au groupe et éviter à tout prix de perturber ceux qui étaient décidés à jouer le jeu. Les résultats des personnes qui ont abandonné l'épreuve n'ont pas été pris en considération dans la synthèse finale. Au demeurant, le pourcentage des futurs candidats ou candidates à la maturité qui, compte tenu de leur performance, s'intéressent à des études de médecine ne devrait pas dépasser 80 %. Si l'on considère un échantillon représentatif d'une classe d'âge, on doit admettre qu'un certain nombre de personnes sont dépassées par le test.

L'on n'a pas davantage tenu compte du dernier sous-test "Travail avec soin et concentration". A la différence des autres, on peut s'entraîner à ce dernier test, et d'ailleurs la brochure d'information recommande d'effectuer cet entraînement à plusieurs reprises. Pour des raisons qui tiennent probablement à un manque de motivation, cette partie du test n'a pas été concluante. Faute d'instructions précises, la majorité des personnes testées n'ont pas rempli correctement le questionnaire relatif à la concentration (on a répondu de haut en bas au lieu de gauche à droite, on a utilisé des cercles ou d'autres graphiques au lieu de traits, effectué des marquages trop faibles, marqué la dernière solution proposée, etc.). Lors d'un contrôle ultérieur, on a pu constater à l'aide du lecteur optique que 50 pour cent des questionnaires avaient été remplis correctement.

La plupart des sous-tests comprennent une série d'items "d'essai" disséminés parmi les items valables, qui ont été testés en vue de la mise au point du futur test réel. Ces items d'essai ne sont pas définis dans la version originale allemande, telle qu'elle est utilisée et publiée. Afin de pouvoir effectuer des comparaisons avec les résultats obtenus en Allemagne, on a éliminé, dans les sous-tests "Reconnaissance de fragments de figures", "Figures tubulaires", "Problèmes quantitatifs et formels" les quatre items qui présentaient la plus faible corrélation avec le résultat du sous-test. Pour la même raison, on a également éliminé, dans le sous-test "Compréhension de textes" un groupe d'exercices comportant 6 items.

## Résultats

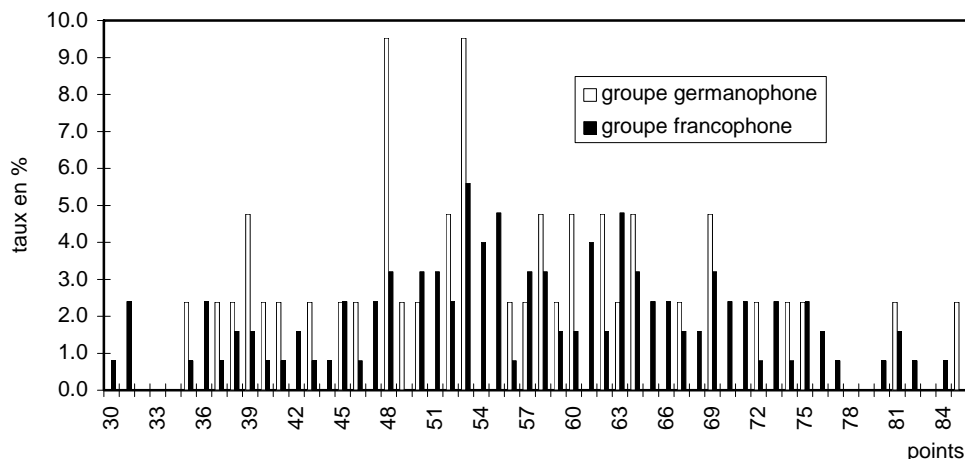
### *Description de l'échantillon*

Au total, 42 lycéennes et lycéens de langue allemande et 125 de langue française (tableau 1) ont effectué le test et tous les 6 sous-tests. La proportion des femmes s'est élevée à 45 %, resp. 56 %. Le nombre correspondant du groupe témoin a été de 23'685 personnes, dont 55 pour cent de femmes. L'âge moyen du groupe fribourgeois s'élevait à 18,4 ans (écart-type de 0,84) contre 19,8 ans pour le groupe témoin allemand (écart-type de 2,80). L'échantillon suisse est d'un an plus jeune que le groupe allemand typique servant de référence.

groupes	nombre [n]	part de femmes [%]	âge	
			moyenne [m]	E.- type [s]
groupe germanophone	42	45	18,4	0,84
groupe francophone	126	56	18,4	0,68
groupe témoin RFA	23685	55	19,8	2,80

Tableau 1: Caractéristiques des personnes testées

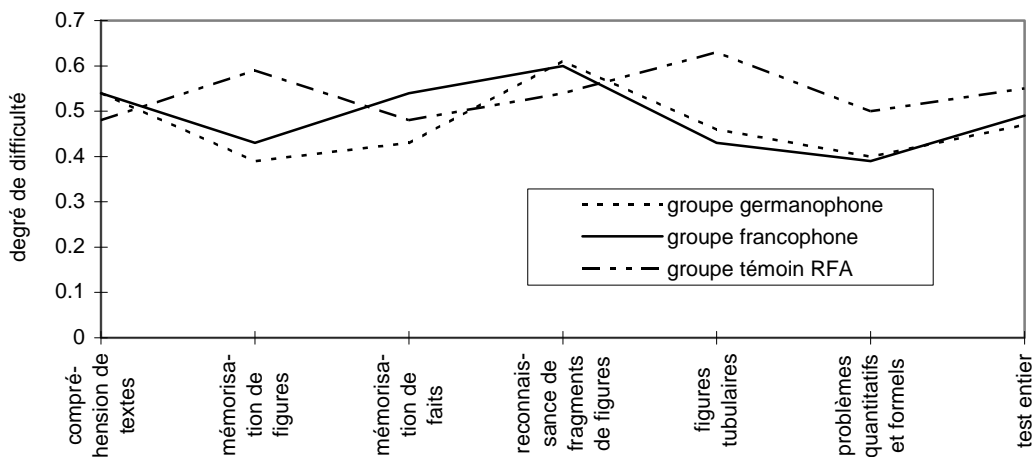
Les personnes testées se sont vu attribuer un point pour chaque exercice coté effectué correctement. Il était possible d'obtenir 118 points au maximum (soit 156 points, moins 18 pour les exercices d'essai disséminés et moins 20 pour le test de concentration). Dans l'épreuve-test, le meilleur résultat obtenu a été de 85 points (graphique 1). Le nombre de points le plus faible attribué est de 30 (pour une valeur moyenne de 56,80 et un écart-type de 12,33). Nous ne disposons pas de valeurs comparatives pour les 6 sous-tests du groupe témoin.



Graphique 1: Répartition des résultats bruts au test en pour cent

## LA DIFFICULTÉ DU TEST

La valeur moyenne du nombre de points obtenus a été de 55,36 points pour le groupe germanophone (écart-type de 12,23) contre 57,28 points pour le groupe francophone (écart de 12,38). L'indice de difficulté s'établit ainsi à 0,47, resp. à 0,49 pour l'épreuve-test. Par comparaison, cet indice est de 0,55 en Allemagne (graphique 2).



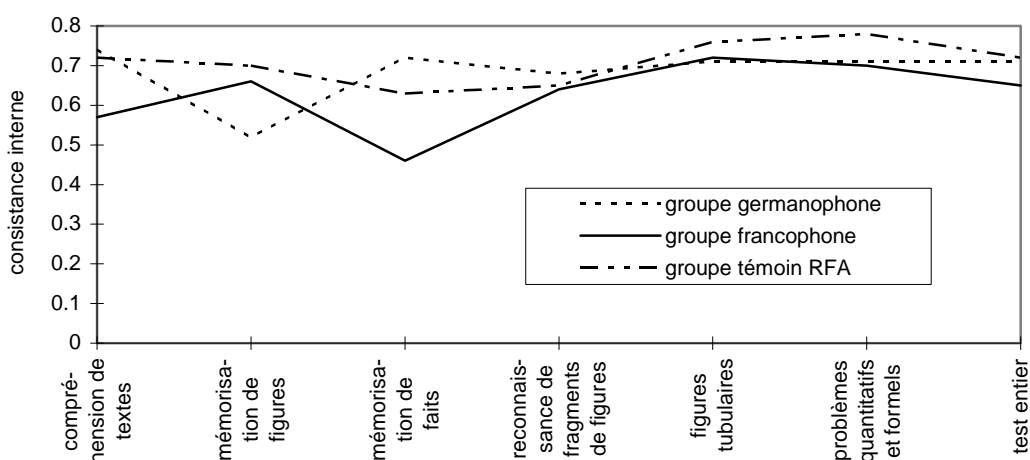
Graphique 2: Degré de difficulté du test et de ses sous-tests

La difficulté moyenne des six sous-tests oscille entre 0,39 et 0,61. Elle se situe ainsi dans l'intervalle de 0,40 à 0,60 permettant ainsi une différenciation optimale. C'est le sous-test "Mémorisation de figures" qui a causé le plus de difficultés au groupe germanophone et le sous-test "Problèmes quantitatifs et formels" au groupe francophone; dans les deux groupes, les personnes testées ont résolu correctement 39 pour cent des exercices cotés. D'autre part, c'est le sous-test "Reconnaissance de fragments de figures" qui a été maîtrisé le plus aisément dans les deux groupes linguistiques (0,61

pour le groupe germanophone; 0,60 pour le groupe francophone). En ce qui concerne ce sous-test (0,54 pour le groupe témoin) et le sous-test "Compréhension de textes" (0,54 pour les groupes germanophone et francophone; 0,48 pour le groupe témoin), les participants suisses ont résolu correctement plus d'exercices que le groupe témoin allemand.

## PRÉCISION DES MESURES

Le "coefficient  $\alpha$  de Cronbach" illustre la qualité des différents sous-tests en tant qu'instruments de mesure. Le graphique 3 indique les valeurs  $\alpha$  de chaque sous-test et la valeur de la médiane correspondante pour l'ensemble de l'épreuve. A l'exception du sous-test "Mémorisation de figures" (0,52), les valeurs  $\alpha$  atteignent la médiane de 0,71 dans le groupe germanophone. Le groupe francophone a obtenu la valeur  $\alpha$  la plus faible de 0,46 pour le sous-test "Mémorisation de faits". Pour l'ensemble de l'épreuve, cette valeur a atteint en moyenne 0,65. Les sous-tests "Figures tubulaires" et "Problèmes quantitatifs et formels" se sont révélés être les instruments de mesure les plus précis pour les participants suisses. Cela correspond aux résultats obtenus en Allemagne.



Graphique 3: Consistance interne du test et de ses sous-tests

## RELATIONS ENTRE LES SOUS-TESTS ET LES NOTES SCOLAIRES

Le tableau 2 répond à la question de savoir dans quelle mesure les champs d'évaluation des sous-tests se recoupent entre eux ou sont au contraire autonomes. Les relations les plus étroites entre les sous-tests se synthétisent dans le sous-test "Problèmes quantitatifs et formels", lequel présente également la plus forte corrélation (0,70) avec l'ensemble du test; c'est aussi un test dont le contenu est très proche du domaine médical.

Les valeurs des corrélations entre les sous-tests ou le test global et les notes scolaires moyennes se situent dans un éventail de 0,06 à 0,23. Cela signifie que la performance globale et les prestations scolaires ont une variance commune maximale de 5,3 pour cent. Il apparaît donc que le test mesure un autre champ que celui exploré par l'école avec ses examens.

Sous-tests	CT	MFI	MFA	RFF	FT	PQF	TOT
Compréhens. de textes (CT)							
Mémoris. de figures (MFI)	0.21						
Mémorisation de faits (MFA)	0.22	0.25					
Rec. de frag. de fig. (RFF)	0.14	0.24	0.03				
Figures tubulaires (FT)	0.26	0.32	0.08	0.30			
Probl. quant. et formels (PQF)	0.35	0.15	0.18	0.38	0.40		
Test entier (TOT)	0.59	0.59	0.47	0.58	0.67	0.70	
Moyenne des notes scolaires	0.19	0.17	0.08	0.22	0.07	0.16	0.23

Tableau 2: Intercorrélations des performances dans les sous-tests ainsi que leurs corrélations avec les notes scolaires

## RELATION ENTRE LE SEXE ET LES RÉSULTATS AU TEST

On considère qu'il y a des différences substantielles dans les performances d'un sous-test dès lors que les valeurs d'efficacité  $\eta$  sont  $> 0,20$ . Des différences entre les moyennes, dans les cas où les valeurs d'efficacité sont supérieures à 0,20, sont significatives même dans un test de t réalisé sur des échantillons indépendants. Dans le sous-test "Problèmes quantitatifs et formels", les hommes du groupe germanophone ont obtenu des résultats supérieurs en moyenne à ceux des filles (tableau 3). En revanche, dans les sous-tests "Mémorisation de figures", "Mémorisation de faits" et "Reconnaissance de fragments de figures", ainsi que pour le résultat d'ensemble du test, ces dernières ont réalisé un résultat moyen significativement meilleur que leurs collègues masculins. Dans le groupe francophone, les filles ont obtenu en moyenne pour les sous-tests "Compréhension de textes", "Mémorisation de figures" et "Mémorisation de faits" un résultat supérieur à celui des jeunes gens de leur groupe. En revanche, ces derniers ont réalisé en moyenne un résultat significativement meilleur que les filles dans les sous-tests "Figures tubulaires" et "Problèmes quantitatifs et for-

mels" et un résultat supérieur dans le sous-test "Reconnaissance de fragments de figures" et pour l'ensemble du test. Si l'on compare les résultats avec ceux du groupe témoin (tableau 4), on obtient une image hétérogène, sans que l'on puisse discerner un avantage de l'un ou de l'autre sexe. Mis à part le sous-test "Compréhension de textes", le groupe de langue française présente des effets analogues à ceux du groupe témoin. Pour le sous-test "Compréhension de textes", les hommes du groupe témoin ont réalisé une valeur moyenne légèrement supérieure, et vice versa pour les femmes du groupe francophone.

sous-tests / test	groupe germanophone					groupe francophone				
	hommes(n=23)		femmes(n=19)		$\eta$	hommes(n=48)		femmes(n=60)		$\eta$
	m	s	m	s		m	s	m	s	
Compréh. de textes	9.04	3.91	10.37	3.15	0,19	9.60	3.04	9.80	2.83	0,04
Mémoris. de figures	6.87	2.67	8.84	2.39	0,37	8.04	3.61	8.90	3.24	0,12
Mémoris. de faits	7.26	2.86	10.26	3.83	0,42	10.52	2.79	11.07	2.99	0,09
Rec. de frag. de fig.	11.35	3.52	13.16	3.24	0,26	12.17	3.71	11.73	2.83	0,07
Figures tubulaires	9.09	3.58	9.11	3.57	0,00	9.92	3.94	7.83	3.43	0,26
Probl. quant et form.	8.26	3.98	7.84	4.36	0,05	9.29	4.26	6.87	3.00	0,31
test entier	51.87	11.52	59.58	12.01	0,32	59.54	14.55	56.20	10.84	0,13

Tableau 3: Relation entre le sexe et les résultats de l'échantillon pour le test entier ainsi que pour les sous-tests

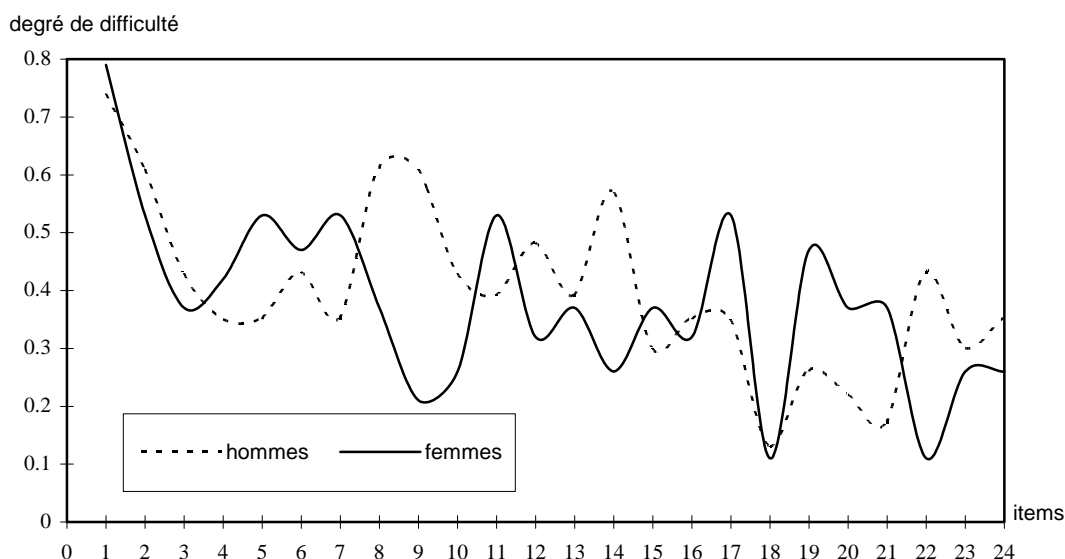
sous-tests	groupe témoin RFA				
	hommes (n=10'707)		femmes (n=12'973)		$\eta$
	m	s	m	s	
Compréh. de textes	9.13	3.68	8.28	3.49	0,12
Mémoris. de figures	11.41	3.70	11.98	3.67	0,08
Mémorisation de faits	9.11	3.38	9.91	3.50	0,12
Rec. de frag. de figures	10.74	3.26	10.88	3.22	0,02
Figures tubulaires	13.43	3.79	11.95	3.92	0,19
Probl. quant. et formels	11.32	4.32	8.73	3.83	0,30
test entier	65.14	6.98*	61.73	6.67*	0,15*

Tableau 4: Relation entre le sexe et les résultats du groupe témoin pour le test entier ainsi que pour les sous-tests (\* = estimation)

Les différences d'ordre statistique concernant les sous-tests qui comportent, chacun, quelques 20 exercices, ne renseignent pas beaucoup sur les caractéristiques des différents items. Les effets particuliers des items peuvent se perdre, selon les cas, au niveau du test ou d'un sous-test (voir Klieme 1991). Un examen détaillé des items a pour but d'identifier les parties du test qui peuvent contribuer à désavantager l'un des deux sexes. Le degré de difficulté, la sélectivité, et la propension à recourir à des techniques de conjecture représentent quelques critères d'items susceptibles de défavoriser l'un ou l'autre groupe.

L'examen des résultats de la présente étude en fonction d'items caractéristiques repose sur une procédure mise au point par Mantel & Haenszel (1959). Compte tenu de la taille réduite de l'échantillon testé, on court cependant le risque que les différences soient dues au hasard. Pour effectuer les calculs l'on a utilisé un programme de Rogers & Hambleton (1994). Au niveau des items, on constate dans le groupe francophone un biais dû aux items (biais d'item) en faveur des filles pour le 1er exercice du sous-test "Compréhension de textes" et pour le 7e exercice du sous-test "Figures tubulaires", et en faveur des garçons pour le 14e exercice du sous-test "Figures tubulaires". Dans le groupe germanophone, l'on a enregistré 4 biais d'item, tous en faveur des hommes, bien que ces derniers aient réalisé en moyenne un score nettement inférieur à celui des filles. Il s'agit en l'occurrence de l'item 13 du sous-test "Mémorisation de faits", de l'item 23 du sous-test "Figures tubulaires", et des items 9 et 22 du sous-test "Problèmes quantitatifs et formels". A titre d'exemple, nous avons mentionné dans la graphique 4 toutes les valeurs de difficultés (valeurs p) des items du sous-test "Problèmes quantitatifs et formels", pour chaque sexe. Nous avons opté pour le diagramme linéaire, parce que ce graphique indique les difficultés croissantes des items et montre aussi à l'évidence qu'on ne saurait parler d'une discrimination systématique de l'un ou de l'autre sexe.





Graphique 4: Relation entre les résultats de chaque item du sous-tests "Problèmes quantitatifs et formels" et le sexe pour le groupe germanophone

Des deux items 9 et 22, qui présentent un biais par rapport au sexe, nous avons choisi le second pour montrer, à titre d'exemple, comment ils ont été étudiés dans le détail. On trouvera dans le tableau 5 le texte de l'item 22 et sa sélectivité par sexe, ainsi que le mode des réponses différent des lycéennes et des lycéens.

<b>Aufgabe</b>				
<i>Eine gewisse Art von Bakterien vermehrt sich in der Weise, dass am ersten Tag aus einem Bakterium zwei, am nächsten aus zwei vier werden usw. Nach 24 Tagen ist das Gefäß, in dem das erste Bakterium angesiedelt wurde, voll.</i>				
<i>Hätte man statt einem vier Bakterien angesiedelt, wann wäre das Gefäß voll geworden?</i>				
réponse	femmes		hommes	
	n	%	n	%
(A) nach 5 Tagen	3	15,8	1	4,3
(B) nach 6 Tagen	5	26,3	5	21,7
(C) nach 18 Tagen	5	26,3	1	4,3
(D) nach 20 Tagen	2	10,5	3	13,0
(E) nach 22 Tagen	2	10,5	10	43,5
sans réponse	2	10,5	3	13,0
sélectivité	0,48		0,34	

Tableau 5: Item 22 du sous-test "Problèmes quantitatifs et formels"

Ni le contenu de l'exercice, ni la solution (E) ne fournissent des éléments qui prouveraient que les femmes du groupe germanophone ont rencontré

des difficultés significativement plus grandes que les hommes pour résoudre cet exercice. Si l'on considère le résultat final, la sélectivité obtenue chez les femmes (0,48) montre, que les meilleurs ont résolu correctement cet exercice, ce qui correspond d'ailleurs au but de l'élaboration d'un item: pouvoir sélectionner les personnes testées d'après leur performance en cours d'exercice. Le contenu de l'item n'autorise aucune interprétation concluante qui permette d'expliquer l'origine de ce biais par rapport au sexe. Au demeurant, on arrive au même résultat pour les autres items dans les deux groupes linguistiques.

## RELATION ENTRE LA LANGUE ET LE RÉSULTAT AU TEST

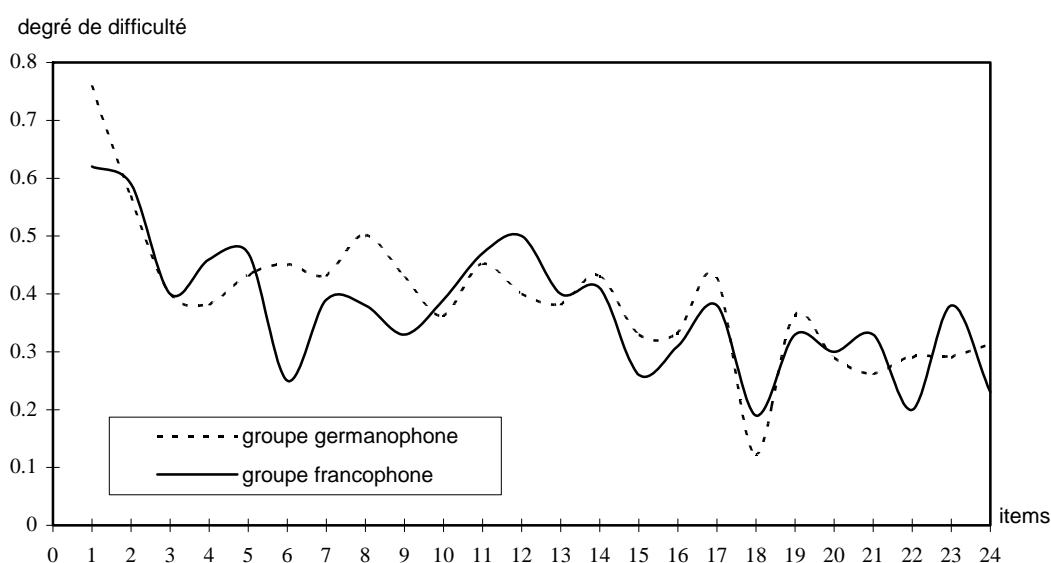
Les différences entre les groupes linguistiques peuvent être mises sur le compte de la traduction ou des performances réalisées. Le design utilisé en l'occurrence ne permet pas de fixer la part à attribuer à chacun de ces deux facteurs.

Il ne s'agit pas ici de savoir quel est le groupe suisse qui a réalisé les meilleurs résultats, mais d'examiner la qualité de la traduction française. Les deux groupes linguistiques s'écartent de manière significative l'un de l'autre uniquement pour le sous-test "Mémorisation de faits" (tableau 6). Dans l'ensemble, le groupe de langue française a obtenu en moyenne un résultat légèrement supérieur à celui de leurs collègues de langue allemande.

	groupe germano- phone (n=42)		groupe franco- phone (n=125)		mesure de dépendance $\eta$
	m	s	m	s	$\eta$
sous-tests					
Compréhension de textes	9.64	3.61	9.63	2.98	0,00
Mémorisation de figures	7.76	2.70	8.55	3.47	0,01
Mémorisation de faits	8.62	3.62	10.75	2.85	0,29
Reconnaissance de fragments de figures	12.17	3.48	11.98	3.19	0,03
Figures tubulaires	9.10	3.53	8.65	3.77	0,05
Problèmes quantit. et formels	8.07	4.11	7.72	3.67	0,04
test entier	55.36	12.23	57.28	12.38	0,07

Tableau 6: Relation entre la langue et le résultat au test

L'analyse des biais d'item entre les groupes linguistiques montre que, pour les sous-tests "Reconnaissance de fragments de figures" et "Mémorisation de figures" - qui sont neutres d'un point de vue linguistique - trois items ont dégagé un biais. Les items 5 (en faveur du groupe francophone) et 22 (en faveur du groupe germanophone) présentent un biais pour le sous-test "Compréhension de textes", qui dépend de la langue. Il en va de même pour l'item 6 (en faveur du groupe germanophone) pour le sous-test "Problèmes quantitatifs et formels". Le graphique 5 sert à illustrer les valeurs de p du sous-test "Problèmes quantitatifs et formels", par langue.



Graphique 5: Relation entre les résultats de chaque item du sous-test "Problèmes quantitatifs et formels" et la langue pour l'échantillon

On enregistre les mêmes degrés de difficulté de l'item pour les deux groupes linguistiques. Le tableau 7 contient des données détaillées concernant l'item 6 du sous-test „Problèmes quantitatifs et formels“, qui présente un biais à mettre sur le compte de la langue. Il semble en effet qu'au sein du groupe francophone l'on ait davantage cherché à deviner qu'à calculer. La solution (D) s'obtient en divisant la capacité maximale d'O<sub>2</sub> (du sportif par 5 l d'O<sub>2</sub> (par minute (2\*4-3)). Cette façon de résoudre le problème (deviner) a dégagé une faible sélectivité (0.12) dans ce groupe linguistique. Le contenu de l'item 6 a été testé par un autre échantillon de personnes de langue française, ainsi que par d'autres experts bilingues que ceux qui ont présidé aux travaux de traduction. Les connaissances ainsi acquises ont démontré qu'il s'est agi en l'occurrence d'une coïncidence statistique due à la taille réduite de l'échantillon. On arrive également au même résultat si l'on étudie les autres items qui présentent un biais par rapport à la langue.

<b>Aufgabe</b>			<b>Devoir</b>		
<p><i>Der menschliche Körper ist in der Lage, für eine kurze Zeit die für die Tätigkeit erforderliche Energie auch dann bereitzustellen, wenn die Sauerstoffaufnahme nicht völlig ausreicht. Es dürfen jedoch insgesamt höchstens 12 l bis 15 l Sauerstoff fehlen ("O<sub>2</sub>-Schuld"). Wenn ein Sportler, der ein O<sub>2</sub>-Aufnahmevermögen von 3 l pro Minute und ein "O<sub>2</sub>-Schuldvermögen" von 14 l hat, eine Tätigkeit erbringt, die 4 l O<sub>2</sub> pro Minute erfordert, dann kann er diese Tätigkeit 14 Minuten durchhalten.</i></p>			<p><i>Le corps humain est capable de mettre à disposition, pour un temps limité, l'énergie nécessaire à une activité même dans les situations où l'oxygénation ne couvre pas entièrement les besoins en O<sub>2</sub>. Toutefois, la quantité d'oxygène manquante ne doit pas totaliser plus de 12 à 15 l ("déficit en O<sub>2</sub>"). Un athlète ayant une capacité d'oxygénation de 3 l par minute, et supportant un déficit en O<sub>2</sub> de 14 l, est capable de soutenir une activité requérant 4 l d'O<sub>2</sub>/minute pendant 14 minutes.</i></p>		
<p><i>Wann wäre die Toleranzgrenze erreicht, wenn der Sauerstoffverbrauch pro Minute auf das doppelte erhöht würde?</i></p>			<p><i>Quand la limite de tolérance serait-elle atteinte si la consommation d'oxygène par minute était doublée?</i></p>		
<b>Antwort</b>	<b>n</b>	<b>%</b>	<b>réponse</b>	<b>n</b>	<b>%</b>
(A) nach 1,7 Minuten	3	7,1	(A) au bout de 1,7 min.	8	6,4
(B) nach 1,8 Minuten	8	19,0	(B) au bout de 1,8 min.	17	13,6
(C) nach 2,5 Minuten	4	9,5	(C) au bout de 2,5 min.	34	27,2
(D) nach 2,8 Minuten	19	45,2	(D) au bout de 2,8 min.	31	24,8
(E) nach 3,5 Minuten	6	14,3	(E) au bout de 3,5 min.	31	24,8
ohne Antwort	2	4,8	sans réponse	4	3,2
Trennschärfe	0,33		sélectivité	0,12	

Tableau 7: Item 6 du sous-test "Problèmes quantitatifs et formels" en langue allemande et française

## RÉSULTAT AU TEST ET AUTRES CRITÈRES

Etant donné la taille réduite de l'échantillon, les autres critères ont été examinés seulement au niveau du test. L'échantillon était composé de personnes appartenant à un groupe d'âge homogène correspondant à une classe d'âge (tableau 8, voir aussi tableau 1). Pour vérifier la relation entre l'âge et le résultat au test, l'échantillon a été réparti en deux groupes: les "moins de dix-neuf ans" et les "plus de dix-huit ans". Dans le groupe germanophone, les "moins de dix-neuf ans" ont obtenu en moyenne moins de points que leurs collègues de même langue, alors que le résultat est inversé chez les francophones..

âge	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
< 19	25	52.84	13.34	69	59.32	12.21
> 18	13	58.85	10.93	37	55.35	13.09
$\eta$	0,23			0,15		

Tableau 8: Relation entre l'âge et le résultat au test

Les personnes testées avaient également la possibilité d'indiquer dans le questionnaire 1 plusieurs domaines scolaires susceptibles de les intéresser. Il en est résulté des groupes trop petits pour les centres d'intérêts pris individuellement. On a dès lors regroupé ces centres d'intérêts particuliers (tableau 9) en domaines d'intérêt plus vastes, à savoir les langues (anciennes ou modernes), les sciences naturelles (biologie, chimie et physique), les mathématiques et les autres domaines (géographie, économie, sport, art, etc.). Il apparaît ainsi que, dans le groupe germanophone, ce sont les élèves qui, en classe, s'intéressent de préférence aux branches des sciences naturelles qui ont obtenu en moyenne le meilleur résultat. Suivent dans l'ordre les groupes "mathématiques", "autres domaines" et "langues". Du côté francophone, le groupe s'intéressant aux mathématiques précède de peu les groupes "sciences naturelles", "autres domaines" et "langues".

domaines d'intérêt	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
langues	16	51.56	10.98	60	57.32	11.50
sciences naturelles	13	57,15	15,93	52	61,50	12,19
mathématiques	18	55,89	12,69	35	62,00	12,00
autres	23	54.09	13.19	78	57.44	12.96

Tableau 9: Relation entre les domaines d'intérêt scolaires et le résultat au test

Les trois candidats au type de maturité B du groupe germanophone ont obtenu en moyenne les meilleurs résultats au test au sein de leur groupe linguistique (tableau 10). Le test de Scheffé, qui permet d'effectuer des comparaisons a posteriori, ne révèle, dans ce groupe linguistique, aucune différence significative entre les moyennes des différents sous-groupes. En revanche, dans le groupe francophone, les candidats à la maturité de type D

ont réalisé des résultats moyens significativement plus faibles que les candidats des types A, B et C. Dans ce groupe linguistique, les candidats à la maturité de type C ont obtenu en moyenne le plus de points.

type de maturité	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
A				13	59.23	11.93
B	3	57.33	4.51	33	57.61	11.10
C	23	55.96	14.15	35	63.26	10.21
D	12	52.25	11.43	17	46.71	10.56
$\eta$	0,15			0,41		

Tableau 10: Relation entre le type de maturité et le résultat au test

type de maturité	groupe francophone			
	A	B	C	D
A				*
B				*
C				*
D				

Test de Scheffé pour le groupe francophone; \* différence significative ( $p < 0,05$ ) entre les moyennes des maturités A, B, C et la moyenne des maturités de type D

Si l'on considère le choix de la branche d'étude (tableau 11) pour laquelle les personnes qui ont pris part à l'épreuve-test pensent opter, on constate que, dans les deux groupes, les indécis ont réalisé en moyenne le meilleur résultat. Suivent, dans le groupe germanophone, ceux qui probablement ne poursuivront pas des études et, dans le groupe francophone, ceux qui savent quelle est leur future branche d'études, à l'exclusion de la médecine. Le test de Scheffé ne fournit aucune différence significative entre les moyennes des différents groupes considérés du point de vue du choix de la branche d'étude.

choix probable des études	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
médecine	1	50.00	0.00	8	55.75	10.01
pas médecine	15	51.40	14.78	30	58.40	13.79
indécis	17	57.88	10.28	54	58.59	11.63
pas d'études	5	56.20	14.97	5	50.00	13.98
mesure de dépendance $\eta$	0,25			0,15		

Tableau 11: Relation entre le choix probable des études et le résultat au test

Les personnes qui ont pris part à l'épreuve-test pouvaient fournir dans le questionnaire 1 des indications sur la formation de leurs parents. Compte tenu de la taille réduite de l'échantillon, les réponses fournies ont été regroupées en formation de type 1 (inconnue ou aucune, école primaire, école secondaire et école professionnelle) et formation de type 2 (maturité, école professionnelle supérieure, école normale, université et doctorat) (tableau 12). Il apparaît ainsi que les jeunes du groupe francophone, dont la mère a reçu une formation de type 2, ont obtenu en moyenne un meilleur résultat que les autres élèves de leur groupe linguistique. Dans le groupe germanophone, on ne détecte aucune influence de la formation du père ou de la mère; il en va de même pour ce qui est de la formation du père dans le groupe francophone.

formation des parents		groupe germanophone			groupe francophone		
		n	m	s	n	m	s
type 1	père	15	54,67	12,29	45	57,69	13,62
type 2	père	21	55,81	13,70	44	57,57	12,05
mesure de dépend. $\eta$		0,04			0,02		
type 1	mère	24	55,00	11,78	54	55,37	13,17
type 2	mère	11	55,82	16,35	34	60,88	11,64
mesure de dépend. $\eta$		0,03			0,18		

Tableau 12: Relation entre la formation des parents et le résultat au test

Diverses raisons ont pu contribuer au fait que les personnes qui ont participé à l'épreuve fribourgeoise se sont senties dépassées par le test ou ont éprouvé des difficultés à bien se concentrer pendant toute la durée de l'épreuve. Dans le groupe francophone, les personnes qui ont bien assimilé le test et travaillé avec concentration ont obtenu en moyenne un meilleur résultat que les autres jeunes de leur groupe. Parmi les autres, l'on n'a pas relevé de différences au niveau des moyennes dans les résultats au test (tableau 13).

impression donnée par le test	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
test plutôt facile	10	56.40	13.07	12	60.00	14.10
test plutôt difficile	32	55.03	12.16	105	57.30	12.30
mesure de dépendance $\eta$	0,05			0,07		
concentration: plutôt bonne	17	55.59	12.17	31	62.10	13.23
concentration: plutôt mauvaise	25	55.20	12.52	88	56.24	11.70
mesure de dépendance $\eta$	0,02			0,21		
exigence trop élevée: plutôt oui	24	53.17	10.66	72	54.18	10.91
exigence trop élevée: plutôt non	16	56.63	13.76	45	62.80	13.30
mesure de dépendance $\eta$	0,15			0,33		
fatigant: plutôt oui	36	55.53	11.92	98	57.64	12.28
fatigant: plutôt non	5	50.40	13.13	18	57.61	14.55
mesure de dépendance $\eta$	0,14			0,03		
peur du test: plutôt oui	1	62.00	0.00	7	60.86	14.25
peur du test: plutôt non	40	55.33	12.46	109	57.30	12.19
mesure de dépendance $\eta$	0,08			0,07		

Tableau 13: Relation entre le sentiment des personnes testées à l'égard du test et le résultat au test

Dans les deux groupes suisses, près de la moitié des personnes a usé de la possibilité de faire un commentaire (tableau 14). Dans les observations des élèves testés, on a critiqué la date peu favorable de l'épreuve, une aération insuffisante des locaux et l'absence de pause. On a également reproché le manque de motivation et douté de la valeur significative des résultats d'une



telle épreuve. En fait, on constate dans les deux groupes linguistiques que les personnes qui ont fait des observations critiques ont obtenu en moyenne moins de points que leurs camarades.

Commentaire	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
oui	23	54.09	12.92	68	55.22	11.83
non	19	56.89	11.49	57	59.74	12.67
mesure de dépendance $\eta$	0,12			0,18		

Tableau 14: Relation entre la présence d'un commentaire sur le questionnaire et le résultat au test

Seulement 32 pour cent du groupe germanophone et 36 pour cent du groupe francophone ont estimé avoir résolu correctement plus de 50 pour cent des exercices (tableau 15). En moyenne, ces pronostics se sont révélés exacts dans les deux groupes linguistiques (51.2 % dans le groupe germanophone et 53.3 % dans le groupe francophone). Dans les deux groupes linguistiques, les personnes de "plus de 49 pour cent" ont obtenu un résultat significativement meilleur que les personnes de "moins de 50 pour cent".

pourcentage de réponses correctes	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
< 50 %	25	53.32	11.55	54	54.11	11.01
> 49 %	12	60.42	12.69	31	62.90	13.11
mesure de dépendance $\eta$	0,26			0,28		

Tableau 15: Relation entre l'estimation du nombre de réponses correctes données et le résultat au test

Seules quelques rares personnes testées (5 dans le groupe germanophone; 6 dans le groupe francophone) ont prévu que leurs performances à l'épreuve du test seraient meilleures que leurs résultats scolaires (tableau 16). Et de fait, ces personnes ont obtenu en moyenne, dans le groupe germanophone, un meilleur résultat (58.00) que leurs camarades du même groupe linguistique (55.17). Le rapport inverse concernant les moyennes scolaires (tableau 17) semble confirmer les données relatives au rapport test-prestations scolaires dans ce groupe. Etant donné qu'une seule personne sur

6 a indiqué sa note scolaire moyenne dans le groupe francophone, il n'est pas possible d'en tirer d'autres conclusions.

performance au test	groupe germanophone résultats au test			groupe francophone résultats au test		
	n	m	s	n	m	s
meilleure que les prestations scolaires	5	58.00	12.81	6	48.50	9.33
moins bonne que les prestations scolaires	35	55.17	12.58	102	57.45	12.01
mesure de dépendance $\eta$	0,08			0,16		

Tableau 16: Relation entre l'estimation des résultats du test en comparaison avec les performances scolaires et le résultat au test

performance au test	groupe germanophone prestations à l'école			groupe francophone prestations à l'école		
	n	m	s	n	m	s
meilleure que les prestations scolaires	3	4.23	0.25	1	4.50	0.00
moins bonne que les prestations scolaires	29	4.88	0.34	76	4.72	0.29
mesure de dépendance $\eta$	0,51			0,12		

Tableau 17: Relation entre l'estimation des résultats du test en comparaison avec les performances scolaires et le résultat au test

Les personnes qui ont exprimé le souhait de recevoir une communication de leur performance ont réalisé un résultat significativement meilleur que celles qui ont renoncé à cette information (tableau 18). On peut en déduire que le groupe qui a souhaité recevoir cette information s'est efforcé d'obtenir le meilleur résultat possible.

résultat au test demandé	groupe germanophone			groupe francophone		
	n	m	s	n	m	s
oui	15	60.07	11.61	52	62.73	10.88

non	27	52.74	11.97	73	53.40	11.96
mesure de dépendance $\eta$	0,29			0,37		

Tableau 18: Relation entre la demande d'un résultat personnel et le résultat au test

## Discussion

L'organisation d'une épreuve comportant un test d'aptitudes devant servir de critère de sélection en vue d'instaurer un *numerus clausus* pour les études de médecine, a pu être menée à bien, comme prévu. Il est possible de résoudre les problèmes liés à une telle organisation avec les structures existantes. Les expériences réalisées durant le déroulement de l'épreuve seront enregistrées dans le manuel des responsables du test, qui en décrit avec précision toutes les étapes. Il semble absolument nécessaire de soumettre les responsables du test à un entraînement. Les locaux ne devraient pas accueillir plus de 100 personnes. Comme le test aura lieu en juillet, il importe de trouver des locaux qui soient bien climatisés.

### LE TEST SATISFAIT AUX CRITÈRES DE QUALITÉ

En dépit d'un manque de motivation et même si les conditions climatiques n'étaient pas optimales, l'on a atteint des critères qualitatifs comparables à ceux obtenus en Allemagne. Pour ce qui concerne la précision des mesures, les groupes suisses s'en tirent légèrement mieux que le groupe témoin allemand dans 2 sous-tests sur 6. Les différences de niveau constatées après l'analyse des données du test s'expliquent surtout par le manque de motivation qui était sous-jacent à l'épreuve. La fiabilité et l'équité du test correspondent, dans les deux groupes linguistiques, aux exigences d'une procédure basée sur le diagnostic psychologique. En particulier, la fiabilité de la valeur globale du test, qui doit servir d'instrument de sélection, correspond, dans les deux groupes linguistiques, au niveau obtenu par le groupe de référence allemand.

### LE TEST PERMET UNE SÉLECTION OPTIMALE

L'attribution de valeurs brutes, les degrés de difficulté et la sélectivité permettent de différencier les candidates et les candidats en fonction de leur performance au test. Par rapport au groupe témoin allemand (55 %), le degré de difficulté a atteint 48 % en moyenne. Compte tenu de la situation qui

règne en Suisse, il faudrait que la différenciation soit telle qu'elle permette aux meilleures des personnes testées, dans une proportion oscillant entre 75 et 85 pour cent, d'accéder aux études. Dans ce domaine, le test devrait donc suffire pour obtenir la sélection souhaitée. On note que 2,3 pour cent des personnes seulement ont obtenu le même nombre de points. Par conséquent, il est possible de déterminer, avec un degré de précision largement suffisant, un nombre de points minimum correspondant à un quota dont le seul critère serait de se maintenir dans les limites des capacités d'accueil.

### **LE TEST NE MESURE PAS LA MÊME CHOSE QUE LES NOTES SCOLAIRES**

En se basant sur les notes scolaires, on a pu également déterminer dans quelle mesure les performances réalisées au cours du test correspondent aux prestations scolaires. La corrélation relativement faible de 0,23 entre la valeur du test et une note globale (indiquée par la personne testée) démontre que le test ne mesure pas tout à fait la même chose que ce que révèlent les notes scolaires. En Allemagne, de telles corrélations sont un peu plus élevées (0,39 - voir Trost et al. 1994). Pour expliquer cette différence entre les notes scolaires et le résultat au test, Trost et al. se réfèrent à l'influence des critères différents d'évaluation qui sont employés pour les notations scolaires. En Suisse, où la sévérité des notes attribuées lors des examens de maturité diffère d'un canton à l'autre et d'une école à l'autre, le problème de la différence de corrélation pourrait s'accroître. A l'inverse, les critères d'appréciation du test sont uniformes et doivent permettre de compenser ces divergences. On relèvera que la validité du test en tant qu'instrument de diagnostic pour le succès des études n'est en rien inférieure aux notes scolaires.

Au demeurant, l'examen des rapports entre les différents sous-tests révèle que chacun de ces derniers, pris isolément, vérifie ses propres critères, et que, partant, aucun des sous-tests n'est superflu parce que faisant double emploi.

L'analyse des rapports entre le test et d'autres critères a démontré qu'il y avait une corrélation importante entre le type de maturité et le résultat au test. Dans le groupe francophone, les candidats à la maturité de type D ont obtenu un résultat moyen significativement inférieur aux personnes qui préparent une maturité de types A à C. On note des résultats comparables au cours des études de médecine. C'est ainsi qu'à la faculté de médecine de l'Université de Berne, entre 44 % et 50 % des étudiants de première année au bénéfice d'une maturité de types A à C échouent aux premiers examens

propédeutiques, alors que l'échec est de 77 % pour les titulaires d'une maturité de type D (Hofer 1992).

### **LE TEST NE DÉFAVORISE PAS LE FEMMES**

C'est à juste titre que l'on exige que les hommes et les femmes aient les mêmes chances d'accéder aux écoles supérieures et que les progrès réalisés sur ce plan ne soient pas remis en cause du fait de l'application du test. Certaines présentations que l'on a fait du test peuvent laisser croire qu'il est plus ou moins discriminatoire à l'égard des femmes. A dire vrai, cette appréciation du test a toujours concerné les résultats réalisés en Allemagne, qui n'ont toutefois pas été confirmés à Fribourg.

Les résultats de l'épreuve ne laissent apparaître aucune différence pour l'un ou l'autre sexe. Qu'il s'agisse du test, des sous-tests ou des items, en aucun cas il a pu être établi l'existence d'une possible discrimination des femmes en Suisse. Seuls 4 items sur 136 dans le groupe germanophone et 3 sur 136 dans le groupe francophone ont présenté un biais en ce qui concerne le sexe. Au niveau des sous-tests, les femmes du groupe germanophone ont obtenu en moyenne de meilleurs résultats que leurs camarades de groupe dans 5 sous-tests sur 6 (dont 3 d'une valeur statistique insignifiante). Dans le groupe francophone, les résultats des sous-tests sont équilibrés. Alors que les femmes ont réalisé un meilleur résultat moyen pour les sous-tests "Compréhension de textes", "Mémorisation de figures" et "Mémorisation de faits", la situation est exactement l'inverse pour les sous-tests "Reconnaissance de fragments de figures", "Figures tubulaires" et "Problèmes quantitatifs et formels". Au niveau du test, les meilleurs résultats ont été obtenus, en moyenne, par les femmes dans le groupe germanophone et par les hommes dans le groupe francophone. Si l'on établit une comparaison avec le groupe témoin allemand, on constate que les résultats du groupe germanophone sont exactement inversés.

Au demeurant, le fait de constater des différences au cours d'un test ne signifie pas automatiquement que ces différences seront reprises dans le critère de sélection. Une fois la décision politique prise, on peut gommer la différence de la valeur moyenne entre les sexes au moyen d'une procédure de correction statistique. Les quotas d'admission peuvent être ajustés de manière précise aux quotas de candidature des deux sexes. L'utilisation de cette procédure suppose, outre la décision politique, la preuve effective que le test d'aptitudes entraîne en Suisse les différences en question.

## **L'ÉGALITÉ DES CHANCES VAUT AUSSI POUR LES GROUPES LINGUISTIQUES**

Le test original a été mis au point en Allemagne. En reprenant un tel test dans une autre culture et une autre langue, il est possible que l'on crée des différences qui influent sur l'égalité des chances. Il est donc nécessaire que deux groupes d'experts indépendants surveillent l'adaptation et la traduction des items de manière à éviter des pertes ou des déformations de leur contenu. Ce n'est qu'après l'exécution du test que l'on peut vérifier, à l'examen des paramètres de contrôle qualitatifs, si l'adaptation a entraîné des lacunes.

6 items sur 136 ont présenté un biais du point de vue linguistique, encore que 3 d'entre eux provenaient de sous-tests indépendants de la langue. Pour les autres items, l'on a vérifié la sélectivité, les degrés de difficultés, le contenu des exercices et la qualité des réponses dans les différents groupes linguistiques. Les divergences détectées ne sont pas systématiques. Pour les sous-tests "Compréhension de textes", "Reconnaissance de fragments de figures", "Figures tubulaires" et "Problèmes quantitatifs et formels", le groupe germanophone a obtenu un résultat moyen légèrement supérieur au groupe francophone. Ce dernier a toutefois réalisé en moyenne, dans les sous-tests de mémoire "Mémorisation de figures" et "Mémorisation de faits", un nombre de points supérieur, qui lui permet d'enregistrer un avantage sur l'ensemble du test par rapport à l'autre groupe linguistique. La traduction soignée et coûteuse du test en langue française a donc donné un résultat qui permet de parler d'une adaptation réussie dans une autre culture.

## **CONCLUSION**

Toutes les conditions d'application d'un test d'aptitudes pour les études de médecine en Suisse sont ainsi réunies, pour peu que les milieux politiques décident de l'introduire.

L'adaptation du test pour la Suisse alémanique et pour la Suisse romande atteint, sur le plan de la fiabilité et de l'équité, une qualité comparable à la version originale allemande.

En se basant sur les valeurs globales du test, l'on obtient un critère d'admission aux études de médecine qui présente une sélectivité suffisante. La difficulté du test pour les deux groupes linguistiques suisses diffère de peu de celle rencontrée par le groupe témoin allemand.

L'épreuve organisée en Suisse n'a pas permis de confirmer des différences des valeurs du test pour ce qui concerne le sexe. Il apparaît donc d'entrée de cause que les chances sont égales pour les hommes et les femmes.

On retrouve la même égalité de chances si l'on compare les deux groupes linguistiques. Il est ainsi établi que la procédure d'adaptation choisie est juste et équitable.

Le test présente l'avantage de vérifier, non pas les connaissances scolaires, mais l'aptitude aux études de médecine.

L'épreuve a permis de confirmer qu'il est possible de satisfaire en Suisse aux conditions structurelles et techniques que requiert ce test.

## Littérature

- Hänsgen, Klaus-Dieter, Hofer, Rainer & Ruefli, Daniel (1995). Der Eignungstest für das Medizinstudium in der Schweiz. Grundlagen, Anwendung und Probleme. Schweizerische Ärztezeitung, 76(37), 1476-1496.
- Hofer, Rainer (1992). Die Beziehung zwischen Maturitätstyp und Erfolg im 1. Propädeutikum an der medizinischen Fakultät der Universität Bern. Unveröffentlichte Studie. Universität Bern: Institut für Aus-, Weiter- und Fortbildung.
- Institut für Test- und Begabungsforschung (Hrsg.) (1990). Test für medizinische Studiengänge. Aktualisierte Originalversion 2. Göttingen: Hogrefe.
- Klieme, Eckhard (1991). Problemstellung: Fairness von TMS-Aufgaben. In G. Trost (Hrsg.) Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 15. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.
- Rogers, H. Jane & Hambleton, Ronald K. (1994). MH: A fortran 77 program to compute the Mantel-Haenszel statistic for detecting differential item functioning. Educational and Psychological Measurement, 54(1), 101-104.
- Trost, Günter (Hrsg.) (1977-1994). Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 1. - 18. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Trost, Günter (Hrsg.) (1994). Test für medizinische Studiengänge (TMS): Studien zur Evaluation. 18. Arbeitsbericht. Bonn: Institut für Test- und Begabungsforschung.
- Zentrum für Testentwicklung und Diagnostik (Hrsg.) (1995). Test-Info. Eignungstest für das Medizinstudium in der Schweiz. Information für die Anmeldung 1995. Universität Fribourg.