

WANG Yongkang  
Université Sun Yat-sen, Chine  
pierrewang@126.com



Le test est étroitement lié à l'évaluation dans de nombreux domaines dont l'éducation en particulier. À partir de la théorie de test de langues, une analyse sur le TFS8 en Chine sera effectuée dans cet article. Sa validité, sa fiabilité et sa difficulté seront notre principale préoccupation au cours de l'analyse. Nous voulons ainsi en étudier les avantages et les inconvénients tout en donnant quelques conseils dans l'optique de son amélioration.

**Mots-clés :** évaluation ; test de langues ; TFS8.

Testing is closely related to evaluation in many areas, education in particular. Based on the language testing theory, an analysis of the TFS8 in China is made with an emphasis on its validity, its reliability and its difficulty. The author examines the advantages, as well as the disadvantages of this test, and proposes some advice for its improvement.

**Key words :** evaluation ; language testing ; TFS8.

测试在许多领域，尤其在教育领域，与评估密切相关。文章依据语言测试理论对中国的法语专八考试进行分析。在分析中将关注它的效度、信度和难度。我们试图找出该考试的长处和不足，并为它的改进提出一些建议。

关键词：评估；语言测试；法语专八。

## Introduction

Aujourd'hui, l'évaluation est sans aucun doute devenue un enjeu social, politique et économique. Elle a pénétré dans de nombreux domaines de notre vie. Dans le domaine de l'éducation, on ne peut même plus parler d'enseignement sans parler d'évaluation. Le test est un aspect essentiel de l'évaluation de l'enseignement. Grâce à l'ouverture de la Chine vers l'extérieur et au développement de son éducation supérieure, il existe en Chine quelque 95<sup>1</sup> écoles supérieures dotées d'un département ou d'une section de langue française. Dans ce contexte, l'évaluation de l'enseignement du français de ces établissements joue un rôle très important. On a ainsi commencé à tester

le niveau de français des étudiants au niveau national à travers le TFS4 et le TFS8 (Test national de français destiné aux étudiants spécialisés en études françaises - niveau IV et niveau VIII). Dans cet article, nous allons analyser le TFS8 pour apporter une certaine contribution à l'amélioration de l'évaluation de l'enseignement du français en Chine.

## 1. Mesure, test et évaluation

On confond souvent les termes « mesure », « test » et « évaluation » en les utilisant comme synonymes. Notre attention aux similarités superficielles entre ces trois termes tend cependant à obscurcir les caractéristiques distinctives de chacun. Il est donc nécessaire de bien les distinguer.

Selon Bachman, « la mesure dans les sciences sociales est le processus de quantification des caractéristiques des personnes, conformément aux procédures et règles explicites. Cette définition comprend trois caractéristiques distinctives: quantification, caractéristiques et règles et procédures explicites. » (1990 : 18-19).

Pour ce qui est du test, d'après Carroll, « un test psychologique ou éducatif est une procédure destinée à obtenir un certain comportement à partir duquel on peut faire des inférences sur certaines caractéristiques d'un individu » (cité par Bachman, 1990 : 20).

En ce qui concerne l'évaluation, pour Weiss, « l'évaluation peut être définie comme la collecte systématique d'informations dans le but de prendre des décisions. » (*ibid.*, p. 22). Un aspect de l'évaluation est la collecte d'informations fiables et pertinentes. Il faut noter que l'évaluation n'implique pas nécessairement le test et que le test n'est pas nécessairement une évaluation. C'est seulement quand les résultats des tests sont utilisés comme base pour prendre une décision que l'évaluation est impliquée.

## 2. Notions importantes de l'évaluation et du test

Nous présentons maintenant quelques notions très importantes de l'évaluation et du test de langues.

### 2.1 Validité

« La validité est la qualité d'une activité d'évaluation qui fait que cette activité mesure exclusivement et exactement ce qu'elle est censée mesurer, ce pour quoi elle a été conçue. » (Tagliante, 2005 : 23) La validité implique deux aspects : 1) ce que l'on mesure exactement ; 2) quel degré l'on atteint dans l'évaluation. Par exemple, si nous voulons mesurer la compétence communicative des élèves à travers un examen, pour avoir une bonne validité, le contenu de l'examen doit spécifiquement concerner la compétence communicative et non d'autres compétences ; de plus, l'examen peut bien tester et distinguer la compétence communicative des élèves. En réalité, les activités de l'enseignement s'organisent souvent en vertu d'un programme d'enseignement rédigé par des personnes ou un service autorisé. La validité de l'évaluation est liée étroitement à ce programme. Des équations mathématiques statistiques pourraient nous

servir à mesurer la validité, nous ne les citons pas, puisque nous ne les utilisons pas dans cet article.

## 2.2 Fiabilité

Ce terme (*reliability* en anglais) concerne la stabilité du résultat d'un test. « La fiabilité est donc une mesure de la précision, la cohérence, la crédibilité ou l'équité des notes résultant de l'administration d'un examen particulier. » (Henning, 2001 : 74) Si on teste un groupe d'élèves à deux ou plusieurs fois par les mêmes items pendant peu de temps, les élèves peuvent obtenir des résultats similaires, c'est-à-dire qu'il existe une cohérence entre les résultats, on peut dire que le test a une très bonne fiabilité.

On utilise le « coefficient de fiabilité » pour voir la corrélation entre les deux résultats du test de mêmes items et aux mêmes élèves. Plus le coefficient est grand, plus la fiabilité est forte. Pour calculer le coefficient de fiabilité, il existe plusieurs formules mathématiques statistiques qu'il n'est pas nécessaire de citer dans cet article. Il y a beaucoup de facteurs qui menacent la fiabilité : quantité d'items, difficulté des épreuves, environnement du test, émotion et état des élèves, émotion et état des correcteurs, changement de correcteurs, critères de notation, etc.

Il faut noter que, dans cet article, nous utilisons les termes de validité et de fiabilité dans leur sens large et non pas dans le sens statistique pour analyser la validité et la fiabilité du TFS8.

## 2.3 Difficulté

Les tests qui sont trop faciles ou trop difficiles pour un groupe de candidats spécifiques présentent souvent une fiabilité faible. La caractéristique la plus importante d'un item à être déterminée avec précision est sa « difficulté ». « La difficulté d'items est déterminée comme la proportion de réponses correctes, signifiée par la lettre  $p$  » (Henning, 1987 : 49). La formule pour la difficulté d'items est :

$$P = \frac{\sum C_r}{N} \text{ (Formule 1)}^2$$

où  $p$  est la difficulté (proportion correcte),  $\sum C_r$  est la somme des réponses correctes et  $N$  est le nombre des candidats.

Nous pouvons simplement utiliser une autre formule :

$$P = \frac{\bar{X}}{X} \text{ (Formule 2)}^3$$

où  $p$  représente la difficulté,  $\bar{X}$  est la moyenne des notes obtenues à l'item et  $X$  est la note totale de l'item.

En ce qui concerne la difficulté, « les items pourraient se diviser en trois classes : items difficiles, items à difficulté moyenne et items faciles » (Shu, 2009 : 72)<sup>4</sup> :

- Pour les items difficiles :  $p$  est entre 0 et 0.3 ;
- Pour les items à difficulté moyenne :  $p$  est entre 0.3 et 0.7 ;
- Pour les items faciles :  $p$  est entre 0.7 et 1.

Dans les examens, il vaut mieux que les items aient une difficulté moyenne pour avoir une bonne validité.

### 3. Analyse du TFS8

#### 3.1 Présentation du TFS8

Conçus selon le *Programme national de l'enseignement supérieur spécialisé du français pour les deux premières années* (1988) et le *Programme national de l'enseignement supérieur spécialisé du français pour les deux dernières années* (1997, ci-après *Programme pour les deux dernières années*), le TFS4 et le TFS8 s'organisent respectivement à partir de 2004 et 2009. Dans cet article, nous allons analyser seulement le TFS8 en nous limitant uniquement au test de 2009 publié dans le *Guide du test national de français destiné aux étudiants spécialisés en études françaises - niveau VIII (TFS8)*, car les autres n'ont pas encore été publiés. D'ailleurs, le seul test serait déjà représentatif, puisque la structure des items reste presque la même. Depuis 2011, la partie I du test a subi une légère modification à propos de laquelle nous exprimerons certaines objections.

Nous citons ci-dessous la structure du test (en 2009 et en 2010) :

Partie	Forme d'items	Détail	Note
I	Dictée	Écrire un texte d'après l'enregistrement.	10 points
II	Compréhension orale	Section A : Lire d'abord 10 questions et les réponses suggérées, écouter ensuite 10 dialogues et choisir enfin la bonne réponse pour chaque question.	5 points
		Section B : Lire d'abord 10 questions et les réponses suggérées, écouter ensuite un texte et choisir enfin la bonne réponse pour chaque question.	5 points
III	Vocabulaire et grammaire	Section A : Lire les phrases et choisir le mot ou l'expression dont le sens est le plus proche de celui de la partie soulignée.	10 points
		Section B : Compléter un texte avec les mots et expressions qui conviennent (des mots ou des expressions sont suggérés pour chaque trou).	10 points
IV	Version	Traduire un texte français en chinois	12.5 points
V	Thème	Traduire des phrases chinoises en français	12.5 points
VI	Compréhension écrite	Lire 4 textes et choisir la bonne réponse à chacune des questions.	20 points
VII	Expression écrite	Lire un message et rédiger un commentaire.	15 points

Tableau 1 – Structure du test en 2009 et en 2010

En 2011, la première partie se modifie un peu : écouter un texte ayant 20 trous et compléter le texte selon l'enregistrement (chaque trou représente un mot ou une expression).

### 3.2 Effet positif et avantages du TFS8

Le TFS8 se déroule à la fin des études universitaires avec pour objectif de contrôler l'apprentissage et l'enseignement du français durant les deux dernières années universitaires et d'obtenir les données nécessaires à l'ajustement possible de l'enseignement. Il s'apparente ainsi à une évaluation sommative. Ce test joue un rôle important pour motiver les étudiants dans leurs études. Il est aussi un phare qui dirige les professeurs dans leur enseignement. D'ailleurs, le Groupe du test de français réalise chaque année une analyse statistique après l'examen et envoie les résultats à chaque section de français. Les données envoyées permettent aux enseignants de voir les lacunes ou faiblesses dans les études de leurs élèves et d'améliorer ainsi l'enseignement dans le futur. Les enseignants feront plus attention au développement des diverses capacités langagières des élèves. Ce test a bien contribué à l'enseignement et à l'apprentissage. Leur amélioration peut se voir par l'augmentation de la note moyenne de tous les candidats chaque année (55.25/100 en 2009, 57.1/100 en 2010 et 58.81/100 en 2011)<sup>5</sup>.

Selon les moyennes précédentes et la formule 2, la difficulté du test de chaque année peut être calculée : 0.5525 en 2009, 0.571 en 2010 et 0.5881 en 2011, ce qui nous permet d'estimer que les tests passés ont tous une difficulté moyenne du point de vue de la totalité d'items.

La structure des items est raisonnable. En revoyant les différentes parties du test, nous trouvons que les items objectifs (partie I, partie II, partie III et partie VI) représentent 60% de la note totale et les items subjectifs 40%. Les questions objectives peuvent garantir une fiabilité forte, mais s'il y en a trop, on ne peut pas tester pleinement les capacités langagières des élèves et il nuit ainsi à la validité. C'est pourquoi nous avançons l'idée que la structure des items est raisonnable.

Ce qui nous préoccupe particulièrement, ce sont les parties I (*Dictée*) et IV (*Version*) : selon nous, il faut que chaque mot présent dans ces deux parties, tant dans l'enregistrement que dans les lignes françaises, vienne du *Vocabulaire* listé dans le *Programme pour les deux dernières années*, sinon, les candidats s'égareront pendant le test et la validité en souffre, car le test manque de cohérence avec le programme. Cependant, comme on présente par avance les nouveaux mots et leurs sens en chinois dans les deux parties, cela atténue ce problème.

Enfin, la section A de la partie II semble adaptée à l'évaluation des compétences des candidats. En effet, dix situations imitant la vie réelle dans les dix dialogues peuvent bien tester la capacité de réaction et la compétence langagière des candidats.

Nous nous proposons ensuite d'analyser les points à améliorer de ce test, qui nous semblent plus significatifs.

### 3.3 Effet négatif et désavantage du TFS8

#### 3.3.1 Effet négatif

D'après le *Rapport*, il existe un effet négatif du TFS : le TFS provoque une pression chez les professeurs car beaucoup d'établissements visent un haut taux de réussite des élèves à ce test. Pour que leurs élèves obtiennent une bonne note et que leur établissement jouissent ainsi d'une certaine réputation, certains professeurs organisent les activités de l'enseignement centrées sur le test et arrêtent même le contenu des manuels déterminés dans leur programme d'enseignement.

Chaque année, après le test, le Groupe du test de français n'annonce pas les résultats des autres écoles à chaque établissement et ne fait pas de comparaison entre eux. On peut dire que la réputation de l'établissement n'en pâtit pas même si les résultats de ses étudiants dans ce test ne sont pas bons. Il nous semblerait donc plus adapté que les enseignants recommandent aux élèves des livres et d'autres dossiers et leur donnent des conseils pour se préparer au TFS au lieu d'arrêter les activités normales de l'enseignement en classe.

#### 3.3.2 Validité du contenu

Le cadre de référence de test dans le *Programme pour les deux dernières années* propose de tester le vocabulaire et la grammaire, la lecture, l'audition, l'expression orale ou la traduction orale, l'écriture, le thème et la version des élèves. Pourtant, les conditions contraintes, l'expression orale ou la traduction orale manquent au TFS8. La validité du contenu de ce test en relation avec la cohérence avec le programme n'est donc pas assurée. Nous avons deux suggestions pour résoudre ce problème : 1) que le Groupe du test de français organise un test d'oral dont le certificat serait indépendant de celui du TFS8. Seuls les étudiants dont la note dépasse un certain nombre de points au TFS8 pourraient y participer ; ou 2) autoriser chaque section de français à tester ses propres élèves ; mais dans ce cas, chaque établissement devrait envoyer la vidéo au Groupe du test de français pour qu'un contrôle soit effectué par prélèvement aléatoire. Le certificat est aussi indépendant et décerné par l'État. Dans le deuxième cas, la formation au test des professeurs est indispensable avant le test oral.

#### 3.3.3 Objections au sujet de la modification de la partie I

Dans un paragraphe plus haut, nous avons parlé de la modification concernant la dictée dans le test à partir de 2011. Notre avis va cependant à l'encontre de ce changement. Voyons d'abord les statistiques de ces trois dernières années : (tableau 2)<sup>6</sup>

Année	Moyenne de la note de chaque partie						
	Partie I	Partie II	Partie III	Partie IV	Partie V	Partie VI	Partie VII
2009	2.26	6.49	8.99	6.68	8.77	12.37	9.68
2010	1.72	6.25	11.37	9.1	7.64	12.83	8.65
2011	5.79	5.5	10.07	7.58	6.89	13.28	9.7

Tableau 2 - Notes obtenues par les étudiants de toute la Chine des années 2009 à 2011<sup>7</sup>

Basée sur la note totale de chaque partie et la formule 2, la difficulté de chaque partie peut être calculée (voir tableau 3) :

Année	Difficulté de chaque partie						
	Partie I	Partie II	Partie III	Partie IV	Partie V	Partie VI	Partie VII
2009	0.226	0.649	0.4495	0.5344	0.7016	0.6185	0.6453
2010	0.172	0.625	0.5685	0.728	0.6112	0.6415	0.5767
2011	0.579	0.55	0.5035	0.6064	0.5512	0.664	0.6467

Tableau 3 - Difficulté de chaque partie calculée à partir des notes obtenues par les étudiants de toute la Chine des années 2009 à 2011

Selon les coefficients de difficulté, nous pouvons voir que la plupart des parties ont une difficulté moyenne. La seule partie difficile en 2009 et en 2010 est la partie I. À cause de la modification, la partie I n'est plus difficile en 2011. L'objectif de la modification est que l'on espère que les élèves peuvent obtenir une note meilleure. Nous contestons la modification pour deux raisons : 1) on demande aux candidats d'écrire un texte dans la dictée au TFS4, compléter les trous d'un texte au TFS8 constitue par conséquent un recul; 2) on doit modifier les critères de la notation au lieu de modifier la forme de la dictée.

En effet, il est plus difficile d'écrire un texte que de compléter les trous d'un texte dans la dictée. Mais la première forme est plus importante que la seconde. Après les études universitaires, les étudiants vont travailler. Au travail, un interprète doit comprendre un discours complet non seulement quelques mots d'un discours. Les étudiants s'entraînent d'après la forme du test avant de le passer. Pour réussir à compléter un texte dans la dictée, ils risquent de faire beaucoup moins d'efforts dans l'entraînement, ce qui nuit à l'amélioration du niveau de la compréhension orale. C'est la raison pour laquelle nous proposons une idée éclectique : en gardant sa forme, on modifie les critères de la notation pour que la partie I ait une difficulté moyenne.

Les critères pour noter la dictée en 2009 et en 2010 sont les suivants :

- pour une faute : - 0.5
- pour 20 fautes différentes : - 10
- une même faute n'est comptée qu'une seule fois

Nous estimons que les critères sont trop sévères. Du point de vue de l'approche communicative, les fautes pour les accents, les majuscules ou la ponctuation, etc. ne sont pas très graves. Nous pensons qu'il vaut mieux écrire un texte dans la dictée, mais le premier critère de la notation devrait être modifié comme

suivant : « à partir de 3 fautes : - 0.5 ». Dans ce cas-là, on peut mieux tester le niveau des élèves et on peut avoir une meilleure validité.

### 3.3.4 Imprudence dans la rédaction du corrigé

Il arrive que des problèmes surviennent dans la rédaction du corrigé. Par exemple, la réponse dans le corrigé à la question 25 (dans la section A de la partie III du test de 2009) n'est pas correcte. Voyons cet item :

25. Atteinte d'une cataracte congénitale, Ping Yali est malvoyante depuis sa naissance.  
A. est mal vue            B. a une mauvaise vue  
C. aveugle                D. a une courte vue

Dans le corrigé, la réponse est C. Dans le dictionnaire *LAROUSSE*, *malvoyant* a le sens «dont l'acuité visuelle est très diminuée», *aveugle* «privé de la vue». Nous croyons que la réponse est B : *avoir une mauvaise vue* peut bien expliquer *être malvoyant*, ce qui est confirmé par des collègues français. Nous recommandons une grande prudence dans la rédaction du corrigé, sinon tant la validité que la fiabilité sont affaiblies.

### 3.3.5 Qu'est-ce que la *compréhension* ?

La formulation des questions liées à la *compréhension* doit se baser sur les nouvelles informations. Les candidats doivent émettre un jugement correct en s'appuyant sur ces nouvelles informations. Dans la compréhension écrite, « ... il ne faut pas poser les questions aux candidats qui connaissent les réponses avant, c'est-à-dire les questions auxquelles les candidats peuvent répondre sans lire les textes » (Li, 1997 : 308). Il en va de même pour la compréhension orale : les questions auxquelles les candidats savent répondre sans écouter l'enregistrement ne sont pas valables. Mais dans le TFS8 de 2009, il existe trois items de ce genre. Les voici :

15. La révolution industrielle a eu lieu (compréhension orale)  
A. en Bretagne  
B. en Europe  
C. en Angleterre  
Sans écouter, les candidats ayant des connaissances en histoire peuvent choisir directement la réponse C (conforme au corrigé).
43. Qui est Aristote ? (compréhension écrite)  
A. Un philosophe grec.            B. Un philosophe français.  
C. Un naturaliste grec.            D. Un botaniste français.

Cet item est comparable au précédent, sans lire le texte, on peut faire son choix. Si on veut tester les connaissances culturelles, on pourrait ajouter une partie spécifique (la littérature, la société et l'histoire de France ; un peu de connaissances en civilisation du monde) dans le test, ce qui est conforme au *Programme pour les deux dernières années*.



61. La sinologie est l'étude (compréhension écrite)
- A. de la culture chinoise.
  - B. de l'histoire de la Chine.
  - C. de la langue, de la culture et de l'histoire de la Chine.
  - D. de la philosophie chinoise.

Celui qui connaît le mot *sinologie* peut aussi choisir sans hésitation la réponse C.

En tout cas, ce genre d'items qui ne testent pas vraiment la *compréhension* et nuisent à la validité du test, ils doivent être évités.

### 3.3.6 Analyse statistique insuffisante

L'analyse statistique après le test est aussi une démarche importante du processus du test. Le Groupe du test de français fait chaque année une analyse sur les résultats du test, mais les données statistiques sont loin d'être suffisantes : on ne voit ni le coefficient de validité, de fiabilité, de difficulté, de discrimination, ni la valeur de variance, de *skewedness* (degré d'asymétrie de la distribution) et de *kurtosis* (degré d'écrasement de la distribution). Afin que le test soit plus scientifique, l'informatique peut toujours servir - un logiciel offre de vastes possibilités. Les théories de l'évaluation et du test sont indispensables pour les auteurs du test et les statisticiens. Hormis les ouvrages cités dans la bibliographie, nous recommandons, à la fin de cet article, quelques autres ouvrages qui traitent de ces théories.

### Conclusion

Dans cet article, nous avons distingué très brièvement les termes *mesure*, *évaluation* et *test*. Nous avons aussi présenté rapidement quelques notions du test de langues. Nous avons également analysé le TFS8 en Chine à l'exemple de celui de 2009 en analysant ses avantages et ses inconvénients. Des conseils ont été donnés afin de contribuer un peu à son amélioration. Le processus complet d'un test compte trois phases : la préparation, l'organisation et l'analyse statistique. Chaque phase a besoin d'un travail minutieux. Il est souhaitable que cette évaluation de l'enseignement supérieur spécialisé du français en Chine gagne autant sur le plan de la rigueur scientifique que sur celui du prestige.

### Bibliographie

Bachman, L. F., 1990. *Fundamental Considerations in Language Testing*. Oxford : Oxford University Press.

Henning, G., 2001. *A Guide to Language Testing : Development, Evaluation and Research*. Beijing : Foreign Language Teaching and Research Press.

Tagliante, C., 2005. *L'évaluation et le Cadre européen commun*. Paris : CLE international / SEJER.

Cao, D.M., (Ed.) 2009. 曹德明 (等编). 全国高等学校法语专业八级考试指南 *quanguo gaodeng xuexiao fayu zhuanye baji kaoshi zhinan (Guide du test national de français destiné aux étudiants spécialisés en études françaises - niveau VIII(TFS8))*. 上海: 上海外语教育出版社 Shanghai : Shanghai waiyu jiaoyu chubanshe.

Cao, D.M., Wang, W.X., (Ed.) 2011. 曹德明, 王文新 编. 中国高校法语专业发展报告 *zhongguo gaoxiao fayu zhuanye fazhan baogao (Rapport sur le développement de l'enseignement supérieur spécialisé du français en Chine)*. 北京: 外语教学与研究出版社 Beijing : Waiyu jiaoxue yu yanjiu chubanshe.

高等学校法语专业基础阶段教学大纲 *gaodeng xuexiao fayu zhuanye jichu jieduan jiaoxue dagang (Programme national de l'enseignement supérieur spécialisé du français pour les deux premières années (1988))*. 北京: 外语教学与研究出版社 Beijing : waiyu jiaoxue yu yanjiu chubanshe.

高等学校法语专业高年级法语教学大纲 *gaodeng xuexiao fayu zhuanye gaonianji fayu jiaoxue dagang (Programme national de l'enseignement supérieur spécialisé du français pour les deux dernières années (1997))*. 北京: 外语教学与研究出版社 Beijing : Waiyu jiaoxue yu yanjiu chubanshe.

Li, X.J., (Ed.) 2001. 李筱菊. 语言测试科学与艺术 *yuyan ceshi kexue yu yishu (The Science and Art of Language Testing)*. 长沙: 湖南教育出版社 Changsha : Hunan jiaoyu chubanshe.

Liu, R.Q., Han, B.C., (Ed.) 2000. 刘润清, 韩宝成 编著. 语言测试和它的方法 *yuyan ceshi he tade fangfa (LANGUAGE TESTING AND ITS METHODS)*. 北京: 外语教学与研究出版社 Beijing : Waiyu jiaoxue yu yanjiu chubanshe.

Shu, Y.X., (Ed.) 2009. 舒运祥. 外语测试的理论与方法 *waiyu ceshi de lilun yu fangfa (LANGUAGE TESTING : Theories and Methods)*. 上海: 世界图书出版公司 Shanghai : Shijie tushu chuban gongsi.

### Quelques autres ouvrages à recommander

Bu, A.H., (Ed.) 2011. 补爱华. 语言测试方法论 *yuyan ceshi fangfalun*. 上海: 上海交通大学出版社 Shanghai : Shanghai jiaotong daxue chubanshe.

Guo, X.H., He, S., Zhao, D.F., (Ed.) 2008. 郭熙汉, 何穗, 赵东方. 教学评价与测量 *jiaoxue pingjia yu celiang*. 武汉: 武汉大学出版社 Wuhan : Wuhan daxue chubanshe.

Wang, Z.Y., (Ed.) 2009. 王振亚. 现代语言测试模型 *xiandai yuyan ceshi moxing (Modern Approaches to Language Testing)*. 保定: 河北大学出版社 Baoding : Hebei daxue chubanshe.

Wu, Z.M., (Ed.) 2003. 吴尊民. 英语测试的理论与实践 *yingyu ceshi de lilun yu shijian*. 北京: 外语教学与研究出版社 Beijing : Waiyu jiaoxue yu yanjiu chubanshe.

Xu, Q., (Ed.) 1992. 徐强. 英语测试的理论与命题实践 *yingyu ceshi de lilun yu mingti shijian*. 合肥: 安徽教育出版社 Hefei : Anhui jiaoyu chubanshe.

Zou, S., (Ed.) 2005. 邹申. 语言测试 *yuyan ceshi (Language Testing)*. 上海: 上海外语教育出版社 Shanghai : Shanghai waiyu jiaoyu chubanshe.

### Notes

<sup>1</sup> D'après le *Rapport sur le développement de l'enseignement supérieur spécialisé du français en Chine*, dit le *Rapport* après.

<sup>2</sup> Henning, p. 49.

<sup>3</sup> Shu Yunxiang, p. 70.

<sup>4</sup> Les chiffres suivants sont cités par Shu Yunxiang, p. 72.

<sup>5</sup> Chiffres relevés dans le *Rapport*, p. 48.

<sup>6</sup> Données envoyées à chaque établissement après le test.

<sup>7</sup> D'après le *Rapport*, les nombres d'étudiants qui ont participé au TFS8 en 2009, 2010 et 2011 sont respectivement de 1 684, 2 094 et 2 357.