

# L'évaluation en FLE

Joies, peines et plaisirs

## Plan du séminaire

- Le contexte de l'évaluation du FLE
- Le problème de la dimensionnalité
- Les effets de l'examineur
- Les instruments de mesure

## Contextes d'évaluation

- 3 contextes principaux :
  - Immigration
  - Scolaire ou académique
  - Professionnel

## Immigration

- Immigrer comme travailleur qualifié
- 33 632 immigrants en 2010 dans cette catégorie.
- Depuis décembre 2011, pour avoir des points pour la maîtrise du français, il faut faire un examen de FLE : TEF/TEFaQ, TCF/TCFQ, DELF/DALF
- Compréhension et expression orale

## Scolaire ou académique

- Depuis la réforme collégiale de 1993 et le renouveau pédagogique, enseignement et évaluation axés sur le développement des compétences
- Les 4 compétences canoniques de l'enseignement des langues sont donc évaluées : expression/compréhension orale/écrite

## Professionnel

- Au Québec, les examens de français des ordres professionnels, examens administrés par l'Office québécois de la langue française
- Les quatre compétences sont évaluées et le degré de difficulté change selon la scolarité des membres de l'ordre : collégiale ou universitaire

## Obstacles et Problèmes

- Il y a 3 problèmes importants qui complexifient l'évaluation de la performance en FLE :
- 1 – La question de la dimensionnalité
- 2 – Les effets de l'examineur (rater effects)
- 3 – Les instruments de mesure utilisés

## La dimensionnalité

- Pour être valide, une évaluation doit, entre autres, avoir une bonne validité de contenu : évaluer ce qu'il faut évaluer et pas autre chose
- Cette unité de contenu = unidimensionnalité
- Mais la question de la dimensionnalité est complexe et sujette à caution

## Exemples de dimensionnalité

- $3+4=x$
- $\log_2 \sqrt{x} + \log_2 \sqrt{3x-11} = 1, x=?$
- Ces deux items semblent évaluer des compétences strictement mathématiques et le premier est l'item concret le plus unidimensionnel qui soit

- Tom hat drei Äpfel. Es gibt einen Apfel zu Sara. Wie viele Äpfel hat Tom nun?
- Cet exemple, comme les précédents, provient d'un manuel allemand. Cet item, mathématiquement aussi simple que le 1<sup>er</sup>, est clairement multidimensionnel

## Le Facteur linguistique

- Beaucoup d'items, peu importe le domaine, nécessite un minimum de compétence linguistique
- Ce qui n'est pas un problème en FLE, puisque nous voulons justement évaluer les compétences linguistiques

- Mais l'expression orale, par exemple, est toujours, au minimum, bidimensionnelle :
- 1 – Le contenu sémantique
- 2 – La forme, le contenu sémiotique
- Prenons deux exemples :

## Logos

- Les philosophes grecs avaient bien saisi cette dualité inhérente au langage
- Nous devons absolument tenir compte de ce dualisme lorsque nous évaluons une performance en expression orale

## Autre aspect - dimensionnalité

- La dimensionnalité est également fonction de l'utilisation de l'évaluation
- Des évaluations diagnostiques, formatives, sommatives et certificatives ont des conceptions différentes de la dimensionnalité

## Conception statistique de la dimensionnalité

- En complément à la conception « intuitive », il est possible d'étudier la dimensionnalité d'un ensemble de données à l'aide d'analyses statistiques
- Ce concept peut s'expliquer ainsi. Supposons l'ensemble de réponses suivant :

Matrice candidat / réponse  
 items dichotomiques : 0 = mauvaise réponse  
 1 = bonne réponse

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
A	0	0	0	0	0	0
B	1	1	1	1	1	1
C	1	0	1	0	1	0
D	0	1	0	1	0	1
E	1	0	1	0	1	0
F	0	1	0	1	0	1
G	1	0	1	0	1	0



- Il faut cependant être très prudent lorsque vient le temps de faire les analyses statistiques
- Plusieurs méthodes et logiciels existent pour analyser ce problème : de l' $\alpha$  de Cronbach jusqu'à l'algorithme Metropolis-Hastings Robbins-Monro raffiné par Cai (2010)
- Ces études, alliées à une analyse de contenu, peuvent néanmoins être très utiles

## Effets de l'examineur

- Dès que l'on n'évalue plus avec des QCM, la subjectivité des examinateurs entre en jeu
- Or, les examinateurs ne peuvent être parfaitement objectifs et corriger exactement de la même manière toutes les évaluations

- Différents facteurs vont influencer le jugement, certains pertinents, d'autres non
- Il faut être capable d'identifier les effets indésirables afin d'en diminuer l'importance
  
- Mais quels sont ces effets?
- Les recherches en ont identifié plusieurs

## Cas concrets



## Principaux effets

- Halo
- Restriction de l'étendue
- Contraste
- Récence
- Proximité
- Ordre
- Caractéristiques socioculturelles

## Pour identifier les effets

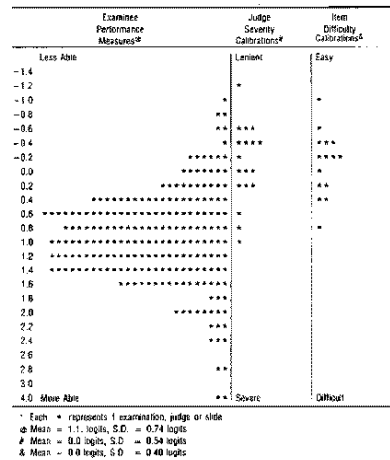
- À moins de cas exceptionnellement flagrants, on ne peut détecter la présence d'effets qu'avec des analyses statistiques
- Pour cela, nous devons utiliser un modèle statistique permettant de tirer des conclusions au niveau individuel, afin de pouvoir attribuer les effets à des examinateurs particuliers

## Modèles statistiques

- Le modèle Rasch à multifacettes n'est pas le seul modèle, mais il est le plus couramment utilisé en évaluation des langues étrangères
- Le principe général est le suivant : on utilise une équation mathématique pour modéliser la compétence des candidats et la difficulté des items.

- La beauté du modèle est que la compétence et la difficulté des items sont ramenés à une même échelle, l'échelle de *logit*
- De plus, on peut aussi modéliser tout autre élément pertinent, par exemple la sévérité d'un examinateur ou le fait de tenir une évaluation le matin ou le soir
- On peut comparer directement tous ces éléments sur une échelle comme celle qui suit

## Exemple de Carte Wright



## Que faire après l'identification?

- Toutes les études ont montré qu'il était impossible d'uniformiser les coefficients de sévérité
- L'objectif est d'augmenter la constance des examinateurs pour que leur coefficient de sévérité soit constant
- Les études tendent à démontrer que les sessions de formation individuelles sont dramatiquement plus efficaces que des formations générales

## Instruments de mesure

- Les instruments de mesure les plus communément utilisés sont les échelles descriptives
- Il faut faire attention à ce que les échelles soient conçues pour être utilisées par des examinateurs

- Les critères et les échelles doivent bien circonscrire les différents concepts (dimensions) évalués
- De la même manière, tous les niveaux de l'échelle doivent être significatifs pour les examinateur, et ce afin de minimiser les risques d'effets de l'examineur

grille-evaluation-ee.pdf - Adobe Reader

File Edit View Window Help

1 / 2 94.3%

Tools Sign Comment

Numéro du candidat + lettre clé : [ ] Centre : ..... Nom de l'examineur : .....

Nom : ..... Date prévue : ..... Prénom : ..... Signature : .....

Prénom : ..... Date effective : ..... Signature : .....

Compétences communicatives

Séances	Ordres	Élémentaire			Intermédiaire		Supérieur	
		04	1	2	3	4	5	6
Section A	Pertinence du questionnement 1	Le questionnement est inapproprié ou ne visant.	Le questionnement est élémentaire et incomplet.	Le questionnement est simple sans demande de précisions.	Le questionnement est satisfaisant : quelques demandes de précisions.	Le questionnement explore l'ensemble de la tâche.	Le questionnement est complet et précis.	Le questionnement est exhaustif et pertinent.
	Prise d'initiative et gestion de l'improvisé 2	Pas ou peu d'initiatives malgré le soutien de l'examineur.	Est souvent sollicité pour engager les échanges et réagit aux interventions.	La prise de parole est présente. Fait reformuler les réponses données.	Fait clarifier les informations ambiguës : souvent trouble par des réponses multiples.	Sollicite spontanément pour demander des précisions : réactions appropriées mais pas toujours immédiates.	Intervient spontanément et face à l'improvisé, réagit avec précision.	Mène avec aisance et efficacité la conversation.
	Qualité des échanges 3	Ne comprend pas les interventions de l'examineur.	De nombreuses ruptures rendent l'échange difficile.	Les échanges sont brefs et non suivis.	Questions et réponses s'enchaînent dans les situations prévisibles.	Les échanges sont suivis même dans les situations imprévues.	Les échanges donnent lieu à une discussion soutenue.	Les échanges sont nets et la discussion enrichie.
Section B	Présentation des faits 4	Les faits sont peu ou pas présentés.	L'exposé est confus ; simple lecture du document.	L'exposé est bref ; les informations présentées sont généralistes.	L'exposé est simple mais clair ; effort de reformulation des informations.	L'exposé est développé et structuré ; les informations sont clairement informées.	L'exposé est clair et complet ; les informations sont présentées avec assurance.	L'exposé est clair, précis et soigné ; toutes les informations sont présentées de façon organisée.
	Qualité de l'argumentation 5	Le discours est décousu ; aucune intention de convaincre.	Le discours est confus ; peu d'arguments présentés.	Le discours est contradictoire et les arguments présentés ne sont pas liés pertinents.	Le discours est cohérent ; les arguments sont justes mais peu développés.	Le discours est clair ; les arguments sont pertinents et liés.	Le discours est bien mené ; les arguments sont convaincants et bien développés.	Le discours est juste et convaincant ; les arguments sont nuancés et appuyés.
	Qualité des échanges 6	Ne comprend pas les interventions de l'examineur.	L'échange est difficile ; réagit à quelques sollicitations.	Les échanges sont directs et brefs ; ne réagit pas toujours aux contre-arguments.	Intervient régulièrement pour justifier son point de vue.	Soutient son point de vue mais peut être trouble par des contre-arguments.	Réagit et défend ses idées avec justesse.	Participe activement à la discussion et défend avec habileté son point de vue.

RTEF-FORT 24 - Version 3 - 1001/2007

MicrosoftPowerP... Atelier Cédame not... Sign'n Mozilla Fire... grille-evaluation-ee...

- Questions?