Université de Moncton, 24 avril 2009

Les modèles de Rasch:
une introduction

Julie Grondin
julie_grondin@uqar.ca

Plan de la présentation

- L'échelle de Guttman (un aperçu)
- La théorie classique des tests (un exemple)
- La mesure, c'est quoi?
- Postulats de base des modèles de Rasch
- Le modèle dichotomique
  - Écriture du modèle
  - Hypothèses et propriété
  - Ajustement des données au modèle
  - Statistiques d'ajustement
  - Critères d'analyse
  - Exemple
  - Comparaison avec la TCT
- Extension au modèle polytomique *Rating Scale*
- Survol de la famille des modèles de Rasch
- Introduction au logiciel Winsteps

Échelle de Guttman

Un aperçu

Échelle de Guttman

- Modèle théorique.
- L'habileté doit se développer de façon monotone croissante.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Sujet 1	1	1	1	1	1	1
Sujet 2	1	1	1	1	1	0
Sujet 3	1	1	1	1	0	0
Sujet 4	1	1	1	0	0	0
Sujet 5	1	1	0	0	0	0
Sujet 6	1	0	0	0	0	0

Échelle de Guttman

- Ne prend pas en compte que les individus peuvent:
  - Deviner;
  - Manquer d'attention, de motivation ou de temps pour terminer;
  - Avoir mal compris le sens de ce qui leur est demandé.
- Ne considère pas que les items
  - Peuvent être ambigus;
  - Ou que la matière peut être nouvellement acquise.
- Etc.

La Théorie Classique des Tests (TCT)

Un exemple

Exemple:

Supposons que l'on doive évaluer l'habileté à skier des individus suivants:

- Enfants
- Adolescents
- Adultes débutants
- Adultes intermédiaires faibles
- Adultes intermédiaires forts
- Adultes avancés
- Athlètes amateurs
- Athlètes aux Jeux Olympiques

Et que l'épreuve utilisée soit la suivante:

Piste école (no 4)
Piste débutante (no 1B ou 15)
Piste familiale (no 1)
Piste intermédiaire no 1 (no 2 ou 7)
Piste intermédiaire no 2 (1D ou 8A)
Piste experte (no 3 ou 5)
Piste extrême (no 13 ou 14)

Carte des pistes du centre de ski Le Relais, Lac Beauport

Supposons les résultats suivants:

		Items								
		1	2	3	4	5	6	7	score total (TCT)	taux de succès (habileté)
Sujets		P.ecole	P. debut	P.familiale	P.inter1	P. inter2	P.experte	P.extreme		
1	at2	1	1	1	1	1	1	1	7	100,0%
2	at1	1	1	1	1	1	1	0	6	85,7%
3 à 5	av 1-3-4	1	1	1	1	1	0	0	5	71,4%
6	av2	1	1	1	0	1	1	0	5	71,4%
7	iA4	0	1	1	0	1	0	1	4	57,1%
8-9	iA 2-3	1	1	1	1	0	0	0	4	57,1%
10	iA1	1	1	0	1	1	0	0	4	57,1%
11 à 13	iF 1-2-3	1	1	1	0	0	0	0	3	42,9%
14	db2	1	1	0	0	0	0	0	2	28,6%
15	db1	0	1	1	0	0	0	0	2	28,6%
16 à 19	do 2 à 5	1	0	0	0	0	0	0	1	14,3%
20	do1	0	1	0	0	0	0	0	1	14,3%
21 à 23	en 1-2-3	0	0	0	0	0	0	0	0	0,0%
score total (TCT)		17	16	13	8	8	3	2		
taux de succès (ind. difficulté)		0,74	0,70	0,57	0,35	0,35	0,13	0,09		
discrimination (6+forts vs 6+faibles)		0,17	0,22	0,26	0,22	0,26	0,13	0,04		

Adaptation des données 2004A d'un projet de recherche subventionné par le Conseil de recherche en sciences humaines du Canada (CRSH) de 2002 à 2005 et de 2005 à 2008.

Analyse des résultats

- L'épreuve est trop difficile pour les enfants.
- L'épreuve est trop facile pour l'athlète olympique.
- Les sujets iA4, iA1, db1 ou do1 présentent des patrons de réponses irréguliers.
- Les adultes obtiennent un score plus élevé que les adolescents.
- Les pistes intermédiaires sont plus difficiles que la piste débutante.

Limites de l'analyse – constat 1

- Les adolescents ont obtenu un score de 14,3%.
- Les adultes débutants, un score de 28,6%.
- Les adultes débutants sont donc plus forts que les adolescents, mais de combien?
- De 14,3%?

Limites de l'analyse – constat 1 (suite)

- L'athlète amateur a obtenu un score de 85,7%.
- L'athlète olympique, un score de 100%.
- L'athlète olympique est donc plus fort que l'athlète amateur, mais de combien?
- De 14,3%?

Limites de l'analyse – constat 2

- Est-ce que la différence de 14,3% reflète bien la quantité d'habileté requise pour qu'un adolescent atteigne le niveau d'habileté d'un adulte débutant?
- Et est-ce que cette quantité d'habileté, 14,3%, est vraiment la même que celle dont a besoin un athlète amateur avant de participer aux Jeux Olympiques?

Limites de l'analyse – constat 2 (suite)

- Fait intéressant, la différence est de 14,3 % partout sur l'échelle.
- Est-ce que la valeur de 14,3% signifie vraiment la même chose sur tout le continuum?

Limites de l'analyse – constat 3

- L'indice de difficulté de la piste école est de 0,74.
- Celle de la piste débutante: 0,70.
- L'indice de difficulté de la piste experte est de 0,13.
- Celle de la piste extrême: 0,09.

Limites de l'analyse – constat 3 (suite)

- Dans les deux cas, on note une différence de 0,04.
- Est-ce que cette différence représente bien la quantité d'habileté requise pour passer de la piste école à la piste débutante, de même que de la piste experte à la piste extrême?

Limites de l'analyse – constat 4

- Est-ce qu'un enfant ou un athlète qui descendent une piste de ski contribuent de la même façon à établir le niveau de difficulté de celle-ci?

Limites de l'analyse – constat 5

- Est-ce que la piste débutante et la piste extrême contribuent de la même façon à établir l'habileté d'un skieur?

Limites de l'analyse – constat 6

- Est-ce que les scores ainsi calculés permettent de prédire facilement la réussite ou l'échec d'un skieur sur une piste?

Limites de l'analyse – constat 7

- Les scores de la TCT, établis comme le taux de réussites des personnes ou des items, dépendent:
  - Du groupe de skieurs évalués;
  - Des pistes de ski utilisées dans l'épreuve.

Conclusion

- Les scores de la TCT permettent:
  - de résumer l'habileté des skieurs à une épreuve ou le niveau de difficulté des items pour un groupe de skieurs;
  - d'ordonner les skieurs selon leur niveau d'habileté ou les pistes selon leur niveau de difficulté;
  - de faire une analyse descriptive des résultats (proportion d'items réussis, proportion de skieurs en échec, etc.)

Conclusion (suite)

- Mais comme les scores de la TCT ne possèdent pas la propriété d'addition, ils ne permettent pas:
  - de contraster les skieurs entre eux;
  - d'évaluer « de combien » un skieur est plus fort qu'un autre;
  - de faire des comparaisons facilement;
  - de dresser un portrait détaillé de l'habileté des skieurs (pistes réussies ou échouées);
  - de faire des prédictions.

La mesure

C'est quoi?

En psychologie, une des définitions de la mesure c'est...

- Assigner des nombres à des objets suivant un ensemble de règles de façon à quantifier un attribut.
- Généralement, les nombres sont ensuite additionnés (par exemple) de façon à produire un score global qui sera utilisé dans des analyses statistiques.

Ce qui n'est pas dit...

- C'est que, pour être mesurable, un attribut doit posséder une structure additive.
- Et donc, que les scores globaux produits, qui ne tiennent pas compte du fait que les nombres assignés aux objets possèdent une structure ordinale (et non pas intervalle ou ratio), ne constituent pas une « mesure » en soit.

Or, la mesure...

- Devrait suivre les mêmes standards que ceux utilisés en physique:
  - Être une abstraction objective d'un objet, d'unités égales;
  - Être reproductible et additive;
  - Être indépendant de l'observateur qui prend la mesure ou des items inclus dans la mesure.

Ainsi, un modèle de mesure utile...

- Doit permettre de faire des inférences, et d'estimer la précision de ces inférences.
- Doit fournir des moyens de détecter et d'évaluer les divergences entre les données recueillies et les paramètres modélisés.
- Doit permettre d'obtenir des résultats invariants.

De plus, un modèle de mesure...

- Doit être sensible à l'acquisition d'habileté et ce, à tous les niveaux d'habileté.
- Doit être capable de nous dire « de combien » une personne est plus habile qu'une autre.

Les modèles de Rasch

Postulats de base

La base du modèle

- Selon Rasch, c'est la différence entre la difficulté et l'habileté qui régit la probabilité qu'un individu réussisse à un item.
- Les idées de base sont que:
  - Une personne d'un haut niveau d'habileté a plus de chances de réussir tous les items;
  - Les items plus faciles ont plus de chance d'être réussis par tous les individus.

La base du modèle (fin)

- Ainsi, les modèles de Rasch sont des modèles mathématiques qui permettent de construire une mesure basée sur une relation probabiliste entre le niveau de difficulté d'un item et le niveau d'habileté d'un individu.
- Ce sont des modèles dits « de trait latent ».

Postulats de base du modèle

1. La performance d'un sujet peut être prédite sur la base de son habileté.
2. La relation entre la performance d'un individu à un item et l'habileté ayant produit cette performance peut être définie à l'aide d'une fonction monotone croissante (FCI).

FCI- Fonction caractéristique de l'item

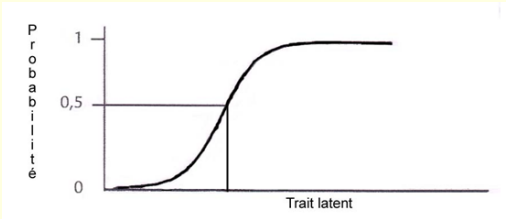
- Fonction monotone croissante.
- Fonction qui relie l'habileté de l'individu à la probabilité de réussir un item.
- Forme de régression logistique basée sur un processus stochastique.

FCI (suite)

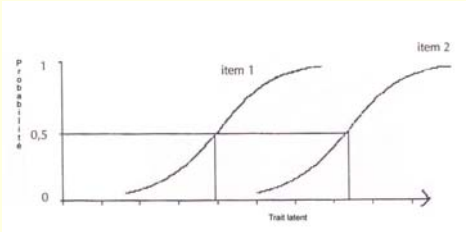
- Transforme les données recueillies en intervalles égaux.
- La représentation graphique de cette fonction s'appelle Courbe caractéristique de l'item (CCI).

CCI- Courbe caractéristique de l'item

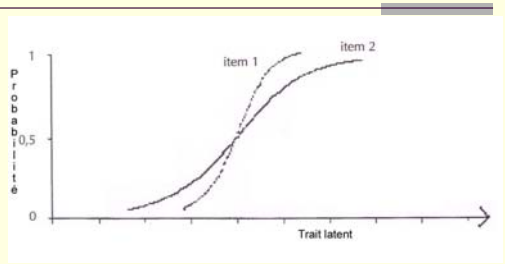
- Courbe résultante de la FCI.



Difficulté d'un item



Discrimination d'un item



Les modèles de Rasch

Le modèle dichotomique

Note importante...

- Un score dichotomique est un score basé sur les valeurs 0 et 1, où 1 à une valeur réelle plus élevée que 0.
- Il ne s'agit pas d'une attribution de valeurs comme dans les bases de données où l'on recode, par exemple, le sexe 0=féminin et 1=masculin.

Mathématiquement...

Le modèle dichotomique s'écrit:

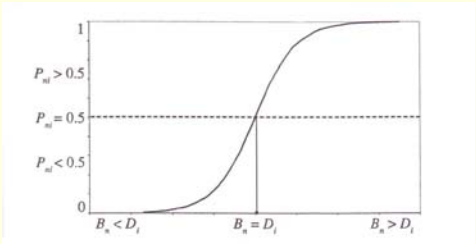
$$Pni = \frac{e\,(Bn - Di)}{1 + e\,(Bn - Di)}$$

En prenant le logarithme naturel du rapport des chances, on obtient :

$$Ln\,[\,Pni / (1- Pni)\,] = Bn - Di$$

Visuellement...

$$Ln\,[\,Pni / (1- Pni)\,] = Bn - Di$$



Le modèle dichotomique...

- Est donc un modèle à deux paramètres:
  - Le sujet;
  - L'item.
- Et il est basé sur les propriétés arithmétiques des échelles d'intervalles.

Échelle logit

- Échelle d'intervalle des modèles de Rasch.
- C'est une échelle en « logit », c'est-à-dire:
  - log (odds)
  - ou logarithme naturel du rapport de chances.
- L'échelle des scores va de – l'infini à + l'infini (en théorie).

Le modèle dichotomique

Hypothèses de base et propriété

Hypothèse 1 – Unidimensionnalité

- Caractère du modèle qui requiert que l'instrument de mesure développé s'attarde sur une seule dimension à la fois.
- Comme en physique, un objet peut posséder plusieurs caractéristiques: longueur, largeur, hauteur, etc. Lors d'une mesure, on s'intéresse à une seule caractéristique à la fois (la longueur par exemple).

Unidimensionnalité (suite)

- Les modèles de Rasch se basent donc sur l'idée que, pour être utile, la mesure doit examiner un seul attribut à la fois sur un continuum de « plus ou moins » de cette caractéristique.
- Même si un objet comporte plusieurs caractéristiques, il est seulement possible d'en mesurer une à la fois.

Unidimensionnalité (fin)

- Signifie qu'un seul facteur dominant, l'habileté d'un individu, permet de déterminer sa probabilité de réussir à un item.
- Condition à remplir avant d'utiliser le modèle.

Hypothèse 2 – Indépendance locale

- Condition selon laquelle les réponses d'un individu sont statistiquement indépendantes.
- C'est-à-dire que la réponse d'un sujet à un item ne permet pas de prédire la réponse qu'il fournira à d'autres items de ce questionnaire.

Indépendance locale (suite)

- Autrement dit, les réponses d'un individu aux items sont seulement déterminées par l'effet conjoint de l'habileté du sujet et du niveau de difficulté de l'item.
- Condition à remplir avant d'utiliser le modèle.
- Propriété fondamentale car elle assure l'indépendance des paramètres.

Indépendance locale (fin)

- Logiquement, l'unidimensionnalité implique l'indépendance locale.
- En effet, l'unidimensionnalité signifie qu'une seule dimension est responsable de la réponse d'un participant à chacun des items: son habileté.
- Ainsi, la réussite ou l'échec de cet individu à d'autres items du questionnaire n'ajoute rien de plus à notre connaissance: indépendance locale.

Propriété – Invariance

- Propriété théorique des modèles de Rasch.
- Propriété selon laquelle les estimations des paramètres d'un modèle sont indépendantes du groupe de sujets ciblés par l'opération de mesure ou du groupe d'items inclus.

Invariance (fin)

- Obtenue lorsque les postulats de base du modèles sont vérifiés.
- Autrement dit, l'invariance est une propriété qui peut être obtenue lorsqu'il y a adéquation entre les données et le modèle.
- Propriété qui assure l'objectivité de la mesure.

Autre caractéristique du modèle: la séparation des paramètres

- Selon les modèles de Rasch, l'estimation d'habileté produite reflète la position d'un individu le long du continuum de « plus ou moins » de la variable mesurée.
- De même, l'estimation de la difficulté reflète la position d'un item le long de ce même continuum.

Séparation des paramètres (fin)

- Ainsi, le rapport de chance qu'un individu réussisse un item est le produit de deux paramètres distincts:
  - un paramètre d'item;
  - et un paramètre pour cet individu.

Le modèle dichotomique

Ajustement des données au modèle

Patron de réponse attendu

- Les modèles de Rasch utilisent une forme probabiliste de l'échelle de Guttman pour établir le patron de réponse attendu dans les données.
- C'est par rapport à ce modèle que les données sont testées pour l'évaluation de l'ajustement entre les données et le modèle.

Ajustement des données au modèle

- De façon classique, c'est le modèle qui est choisi en fonction des données à analyser: les données sont fixes et il faut choisir le modèle qui s'y ajuste le mieux.
- Pour Rasch, le modèle défini la mesure et ce sont les données qui doivent rencontrer les conditions de base du modèle.

Ajustement des données au modèle (fin)

- En fait, la relation entre les données et le modèle est plus symétrique: les sources de discordances peuvent provenir du modèle comme des données.
- Lorsqu'il y a discordance, il faut vérifier si ce sont les données qui s'ajustent mal au modèle ou si c'est le modèle choisit qui n'est pas le bon. (P. ex.: modèle unidimensionnel pour des données à facettes multiples.)

Lorsque le modèle est adéquat...

- Il permet de faire des prédictions.
- Il permet d'obtenir des résultats invariants, c'est-à-dire indépendants des items et des sujets inclus.
- Il fournit des estimés pour chaque item et chaque individu de façon séparée.

Le modèle dichotomique

Calcul de l'ajustement

Estimation des paramètres

- Méthode de maximum de vraisemblance
- Maximiser la vraisemblance (produit des probabilités) c'est:
  - chercher les valeurs Bn et Di qui minimisent la différence entre les scores observés et les scores estimés par le modèle.

Matrice des résidus

- Matrice des différences entre les scores observés et les scores estimés par le modèle.
- La variance (écart-type) n'est pas la même pour chacune des valeurs de la matrice.
- L'interprétation de ces résidus dépend donc des valeurs modélisées.
- La somme des différences sur chacune des colonnes ou des rangées de la matrice est nulle.

Matrice des résidus standardisés

- Matrice des résidus centrée et réduite.
- Permet de contrôler les différences de variance (écart-type).
- Ne permet pas de régler les problèmes de somme sur chacune des personnes ou des items.

Matrice des résidus standardisés, mis au carré

- Permet de faire la somme des éléments sur chacune des rangées ou des colonnes.
- N'annule pas le pouvoir de détecter les réponses inattendues de la matrice.
- La distribution peut s'apparenter à celle d'un khi-carré à un degré de liberté.
- Permet d'examiner l'adéquation entre les données et le modèle.

Le modèle dichotomique

Évaluation de l'ajustement des données au modèle

Évaluation de l'ajustement

- Consiste à comparer les valeurs estimées par le modèle aux valeurs observées.
- Permet:
  - d'identifier les données qui ne rencontrent pas les conditions de base du modèle (qui ne s'ajustent pas bien au modèle);
  - d'examiner les caractéristiques de ces données ;
  - et de vérifier de quelle façon elles contreviennent à la mesure.

Ajustement général

- Correspond à la somme de tous les éléments de la matrice des résidus standardisés mis au carré.
- Lorsque la valeur trouvée est grande, c'est que les items ou les sujets ne s'ajustent pas bien au modèle.

Ajustement général (fin)

- Généralement, les problèmes proviennent de réponses inattendues ou incohérentes.
- Peut signifier que les conditions de base du modèle ne sont pas respectées.

Ajustement pour les items

- Somme des éléments de la matrice des résidus standardisés mis au carré sur chacune des colonnes.

Ajustement pour les personnes

- Somme des éléments de la matrice des résidus standardisés mis au carré sur chacune des rangées.

Mauvais ajustement

- Lorsqu'un sujet ou un item ne s'ajuste pas bien au modèle, il est souvent préférable de le retirer car il participe à établir la mesure des sujets et l'estimation des difficultés des autres items.
- Il pourrait donc fausser l'estimation des paramètres.

Le modèle dichotomique

Statistiques d'ajustement

Indices outfit

- *outfit = outlier sensitive mean square residual goodness of fit statistic*
- Calculés à partir de la matrice des résidus standardisés mis au carré
- Version non pondérée (*unweighted*) des statistiques d'ajustement
- Très sensible aux données aberrantes
- Met l'accent sur les réponses inattendues qui s'éloignent du score de la personne

Outfit (fin)

- Permet de déceler:
  - des erreurs d'inattention
  - des réponses réussies de façon inattendue à des items difficiles par des sujets de faible habileté
  - des mauvaises réponses fournies de façon inattendue par des sujets possédant un niveau d'habileté élevé à des items plus faciles
- Les problèmes signalés par cet indice sont généralement simples à diagnostiquer et à corriger.

Indices infit

- *outfit = outlier sensitive mean square residual goodness of fit statistic*
- Correspond à la version pondérée (*weighted*) des statistiques d'ajustement
- Chacun des éléments de la matrice standardisée mise au carrée est divisé par la fonction d'information.
- L'effet de cette pondération est ensuite annulé en divisant la somme obtenue sur chacune des colonnes par la somme des coefficients (poids) utilisés.

Infit (fin)

- Meilleur choix dans l'étude de l'ajustement des données au modèle.
- Met l'accent sur les réponses inattendues qui sont près du score de la personne.
- Les problèmes décelés grâce à cet indice sont généralement difficiles à diagnostiquer et à corriger. Ils présentent donc un risque plus grand pour la mesure.

Remarque

- Les indices d'ajustement *infit* et *outfit* sont calculés pour :
  - l'ajustement des données au modèle;
  - l'ajustement des items;
  - l'ajustement des personnes.

Le modèle dichotomique

Les autres indices statistiques

Notes sur les indices *infit* et *outfit*

- Peuvent être interprétés comme des statistiques ayant une distribution khi-carré.
- Cependant, pour deux degrés de liberté différents, la valeur de la région critique associée sera également différente.
- Ainsi, il devient impossible de déterminer une valeur unique comme point de référence afin de juger de la qualité de l'ajustement entre les données et le modèle.

Solution 1 – indices *mean square*

- Consiste à diviser les indices *infit* et *outfit* par leur nombre de degrés de liberté.
- Les statistiques sont ainsi transformées en carré moyen (« *mean square* » ou mnsq).
- La valeur attendue d'un carré moyen est de 1, son étendue est de 0 à l'infini.
- Cette statistique n'est pas donc pas symétrique par rapport à la valeur attendue.

Indices *mean square* (fin)

- Par conséquent, si la référence choisie pour juger de la qualité de l'ajustement est symétrique, le taux d'erreur de Type 1 sera différent pour chacune des queues de la distribution.
- Certains problèmes d'ajustement importants ne sont pas détectés par les statistiques des carrés moyens: elles ne décèlent pas le manque d'invariance dans les estimations des paramètres.

Solution 2 – indices standardisés

- Constitue la transformation la plus courante.
- Transforme le carré moyen en racine cubique.
- Permet de convertir le carré moyen en une statistique qui s'apparente à la statistique t de Student.
- Est appelé l'indice d'ajustement standardisé (« *standardized fit index* » ou zstd).

Indices standardisés (fin)

- Permet de développer une valeur de référence possédant un taux d'erreur de Type 1 similaire pour chacune des queues de la distribution, de même que pour plusieurs conditions différentes.
- Plus stable que mnsq lorsque la taille de l'échantillon varie.

Erreur type sur l'estimation

- Le modèle fournit également une estimation de l'erreur type sur chacun des estimés.
- Permet d'évaluer la précision des mesures fournies par le modèle.
- Peut être utilisée afin de décrire une étendue (ou un intervalle de confiance) à l'intérieur de laquelle la « vraie » habileté des personnes ou le « vrai » niveau de difficulté des items se trouve.

Erreur type moyenne

- « *mean square measurement error* » ou mse
- Permet d'évaluer la « fidélité » de l'estimation.
- Consiste à mettre l'erreur type de chacun des estimés au carré, de les additionner et de diviser cette somme par le nombre d'éléments additionnés.
- Donne une précision plus « fidèle » que l'erreur type d'une personne moyenne.

Indice de « fidélité » (*reliability*)

- Tout d'abord, le modèle calcule l'erreur type moyenne (mse).
- Ensuite, cette valeur est soustraite de la variance réelle (observée) du groupe de personnes.
- Consiste donc en une variance ajustée à l'erreur de mesure et représente la « vraie » variance des mesures.
- Permet de déterminer jusqu'à quel point les estimations produites par le modèle sont influencées ou non par des erreurs de mesure.

Indice de séparation

- Calculé comme le rapport entre la variance ajustée (« vraie ») et la variance observée.
- Représente la proportion de la variance qui n'est pas due à l'erreur de mesure.
- Permet d'évaluer la répartition des personnes ou des items (séparation) sur le continuum.
- Homologue à l'indice de cohérence interne KR-20.

Nombre de niveaux de séparation (strata)

- Nombre de niveaux de performance statistiquement différents pouvant être identifiés dans les données.
- Nb de niveaux = Hi = (4 x Gi + 1) / 3
- où Gi correspond à l'indice de séparation.

Analyse graphique du continuum

- Si, en aucun point sur le continuum, on ne retrouve d'items très rapprochés, alors les réponses des participants seraient influencées par une seule dimension, celle mesurée par le test.

Analyse graphique des catégories de réponse

- Permet de vérifier si la courbe de probabilités de chacune des réponses possède un sommet distinct et si les courbes apparaissent comme une suite de collines également distantes.

Le modèle dichotomique

Critères d'analyse

Critères d'analyse

- Les critères présentés sont ceux proposés par Linacre (2004).
- Ce sont des recommandations générales, des pistes d'analyses, plus que des règles strictes à appliquer.

Critères d'analyse

- Tous les items servent à mesurer une même dimension.
- Chacune des catégories de réponse possède au moins dix observations.
- Les observations sont distribuées de façon uniforme dans les différentes catégories de réponses.

Critères d'analyse (suite)

- Ajustement des données au modèle:
  - Les *infit* et *oufit* standardisés sont dans l'intervalle (-2, 2).
- Même chose pour l'ajustement des items et celui pour les personnes.
- Le degré de corrélation entre la mesure estimée par le modèle pour les personnes et la réponse observée est positive (>0). De même pour les items.

Critères d'analyse (suite)

- Vérifier l'erreur type sur l'estimation.
- De même que l'erreur type moyenne.
- L'indice de « fidélité » devrait être d'au moins 0,8.
- L'indice de séparation correspondant devrait alors être plus grand que 2.
- Vérifier le nombre de niveaux de séparation (strata).

Le modèle dichotomique

Un exemple

Revenons à notre exemple...

- Enfants
- Adolescents
- Adultes débutants
- Adultes intermédiaires faibles
- Adultes intermédiaires forts
- Adultes avancés
- Athlètes amateurs
- Athlètes aux Jeux Olympiques

- Piste école
- Piste débutant
- Piste familiale
- Piste intermédiaire no 1
- Piste intermédiaire no 2
- Piste experte
- Piste extrême

Ajustement des personnes

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	person
1	7	7	5.41	1.95	MAXIMUM ESTIMATED MEASURE							at1
2	6	7	3.71	1.34	.54	-.8	.19	2.3	.61	100.0	87.7	at1
3	5	7	2.07	1.26	.24	-1.2	.14	.6	.83	100.0	88.8	av4
4	5	7	2.07	1.26	.24	-1.2	.14	.6	.83	100.0	88.8	av5
5	5	7	2.07	1.26	1.59	1.2	1.05	1.1	.54	71.4	88.8	av2
6	5	7	2.07	1.26	.24	-1.2	.14	.6	.83	100.0	88.8	av1
7	4	7	.61	1.17	3.36	2.8	19.90	3.1	.01	87.1	82.5	iA4
8	4	7	.61	1.17	.71	-.4	.32	.1	.82	85.7	82.5	iA3
9	4	7	.61	1.17	.71	-.4	.32	.1	.82	85.7	82.5	iA2
10	4	7	.61	1.17	2.51	1.9	11.64	.9	.53	85.7	82.5	iA1
11	3	7	-.76	1.20	.27	-1.2	.16	.0	.88	100.0	87.4	iF3
12	3	7	-.76	1.20	.27	-1.2	.16	.0	.88	100.0	87.4	iF2
13	3	7	-.76	1.20	.27	-1.2	.16	.0	.88	100.0	87.4	iF1
14	2	7	-2.16	1.19	.37	-1.1	.19	1.0	.78	100.0	86.2	dn2
15	2	7	-2.16	1.19	1.62	1.0	.84	1.3	.86	71.4	86.2	dn1
16	1	7	-3.65	1.27	.64	-.6	.25	2.5	.54	100.0	85.4	dn5
17	1	7	-3.60	1.27	.64	-.6	.25	2.5	.54	100.0	85.4	dn4
18	1	7	-3.60	1.27	.64	-.6	.25	2.5	.54	100.0	85.4	dn3
19	1	7	-3.60	1.27	.64	-.6	.25	2.5	.54	100.0	85.4	dn2
20	1	7	-3.60	1.27	1.06	.3	.41	2.5	.46	71.4	85.4	dn1
21	0	7	-5.19	1.91	MINIMUM ESTIMATED MEASURE							en3
22	0	7	-5.19	1.91	MINIMUM ESTIMATED MEASURE							en2
23	0	7	-5.19	1.91	MINIMUM ESTIMATED MEASURE							en1
MEAN	2.9	7.0	-.88	1.55	.85	-.3	.89	1.3		91.0	86.0	
S.D.	2.0	.0	2.95	.27	.79	1.1	2.16	1.1		13.3	2.2	

Ajustement des items

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	item
1	16	19	-3.62	.76	1.64	1.8	4.16	1.6	.82	84.2	83.8	1. Piste école
2	15	19	-3.09	.72	.69	-1.3	.28	.3	.76	94.7	85.1	2. Piste débutant
3	12	19	-1.62	.69	.83	-.9	.74	.1	.79	89.5	88.7	3. Piste familiale (facile)
4	7	19	.88	.66	.91	-.8	.83	-.2	.79	89.5	91.2	4. Piste intermédiaire 1
5	7	19	.88	.66	.85	-1.5	.31	-.6	.79	89.5	91.2	5. Piste intermédiaire 2
6	2	19	3.11	.86	.66	-.6	.26	.9	.61	94.7	90.4	6. Piste experte
7	1	19	4.06	1.11	11.44	.9	11.72	2.0	.41	94.7	94.5	7. Piste extreme
MEAN	8.6	19.0	.00	.78	.93	-.2	11.14	.6		91.0	86.0	
S.D.	5.5	.0	2.73	.15	.40	1.1	11.32	.9		5.7	4.6	

Indices de fidélité et de séparation pour les personnes

SUMMARY OF 19 MEASURED (NON-EXTREME) persons

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	3.2	7.0	-.53	1.23	.85	-.3	.89	1.3
S.D.	1.6	.0	2.33	.05	.79	1.1	2.16	1.1
MAX.	6.0	7.0	3.71	1.34	3.36	2.8	9.90	3.1
MIN.	1.0	7.0	-3.60	1.17	.24	-1.2	.14	.0

REAL RMSE	1.37	ADJ.SD	1.88	SEPARATION	1.37	person RELIABILITY .65
MODEL RMSE	1.23	ADJ.SD	1.97	SEPARATION	1.60	person RELIABILITY .72
S.E. OF person MEAN = .16						

MAXIMUM EXTREME SCORE: 1 persons
MINIMUM EXTREME SCORE: 3 persons

SUMMARY OF 23 MEASURED (EXTREME AND NON-EXTREME) persons

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	2.9	7.0	-.88	1.55				
S.D.	2.0	.0	2.95	.27				
MAX.	7.0	7.0	5.41	1.95				
MIN.	.0	7.0	-5.19	1.17				

REAL RMSE	1.48	ADJ.SD	2.55	SEPARATION	1.72	person RELIABILITY .75
MODEL RMSE	1.38	ADJ.SD	2.61	SEPARATION	1.90	person RELIABILITY .78
S.E. OF person MEAN = .63						

person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) person RAW SCORE RELIABILITY = .80

Indices de fidélité et de séparation pour les items

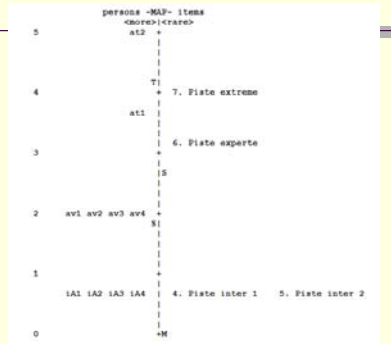
SUMMARY OF 7 MEASURED (NON-EXTREME) items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	8.6	19.0	.00	.78	.93	-.2	1.14	.6
S.D.	5.5	.0	2.73	.15	.40	1.1	1.32	.9
MAX.	16.0	19.0	4.06	1.11	1.64	1.8	4.16	2.0
MIN.	1.0	19.0	-3.62	.66	.55	-1.5	.26	-.6

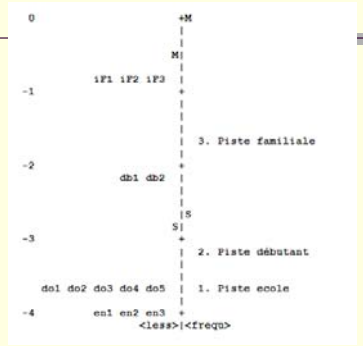
REAL RMSE	.87	ADJ.SD	2.59	SEPARATION	2.98	item RELIABILITY .90
MODEL RMSE	.79	ADJ.SD	2.61	SEPARATION	3.30	item RELIABILITY .92
S.E. OF item MEAN = 1.11						

UMEAN=.000 USCALE=1.000
item RAW SCORE-TO-MEASURE CORRELATION = -1.00
133 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 75.73

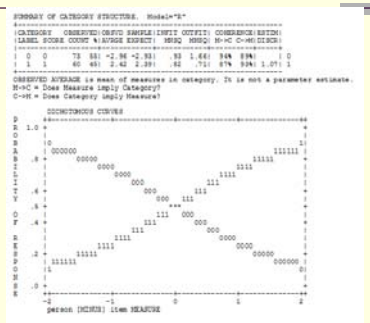
Analyse du continuum



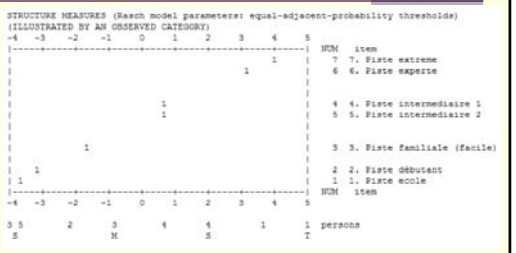
Analyse du continuum (fin)



Analyse des catégories de réponse



Analyse de la structure des catégories de réponse



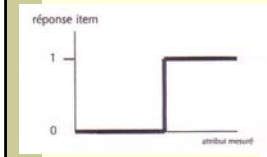
Le modèle dichotomique

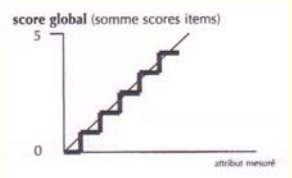
Comparaison avec la TCT

Lien avec la TCT...

- Les modèles de Rasch utilisent le score de la TCT comme point de départ pour estimer la probabilité de répondre correctement à un item.
- Ils transforment ce score brut en une échelle d'intervalle.
- L'échelle (logit) ainsi créée est la même pour les sujets, les items (et les autres paramètres estimés par le modèle).

Les scores de la TCT...

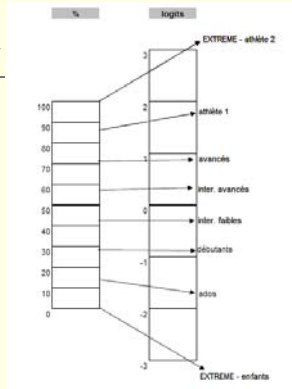




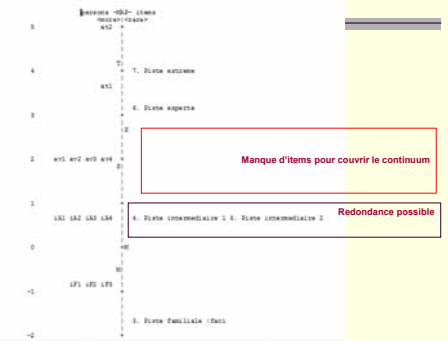
Transfert vers Rasch

- Les scores bruts de la TCT sont convertis en rapport de chance: taux de réussite / taux d'échec. (P. ex.: 60% / 40%)
- Le logarithme naturel est ensuite appliqué à ce rapport de chance, ce qui permet de produire une mesure linéaire et d'obtenir des intervalles égaux.

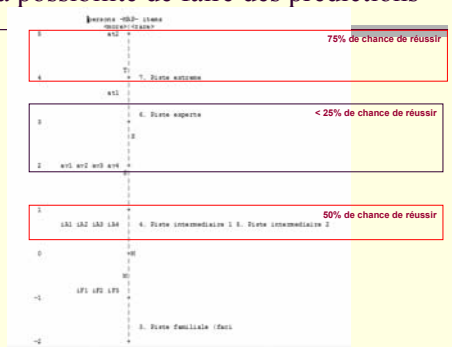
TCT vs Rasch



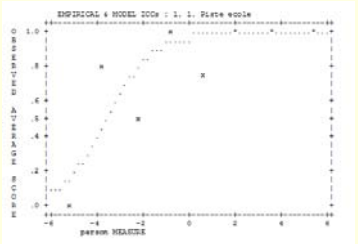
L'analyse du continuum de Rasch



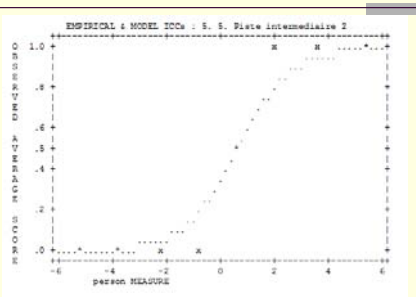
La possibilité de faire des prédictions



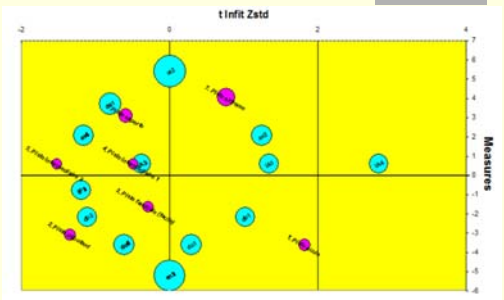
L'analyse des CCI



Analyse des CCI (suite)



Analyse graphique des différents indices statistiques



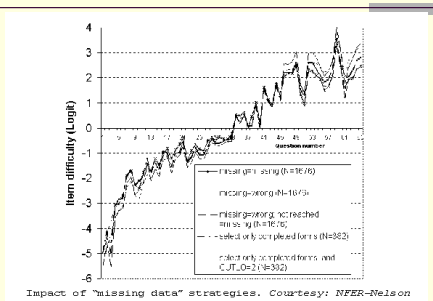
Autre avantage

- Dans notre exemple (évaluation des skieurs), qu'arrive-t-il si un skieur a été malade et n'a pas pu descendre la piste 4 (interm. 1) par exemple?
- Est-ce qu'on lui donne 0?
- La TCT ne nous permet pas de gérer cette situation facilement.

Autre avantage (suite)

- Par contre, les modèles dits de « trait latent », comme ceux de Rasch, sont bien adaptés aux situations comprenant des données manquantes.
- Une étude a testé 4 stratégies différentes pour traiter les données manquantes.
- Leurs résultats montrent que la calibration des items est restée essentiellement équivalente.

Autre avantage (suite)



Impact of "missing data" strategies. *Courtesy: NFER-Nelson*

http://www.rasch.org/rmt/rmt142h.htm

Le modèle dichotomique

FIN

Comment vous sentez-vous?

1. Super!	2. Hein? Je pense que je me suis perdu(e) en cours de route.	3. C'était dur, mais j'y suis arrivé(e).	4. Je ne pense pas que je vais y arriver.	5. Je sens que je vais être malade.

Les catégories de réponses

Les échelles de type Likert

Échelles de type Likert

- Choix de réponses dont la signification serait à peu près la même pour tout le monde.
- Catégories de réponses qui suivent une certaine gradation (ordonnées).
- Généralement, chaque catégorie possède une étiquette ainsi qu'une valeur numérique.
- Permettent de se faire une idée de l'attitude ou de l'opinion des gens.

Échelles de Likert – constat 1

- Ces catégories de réponses sont généralement ordonnées, mais les intervalles entre chacune des catégories sont inégaux.

Échelles de Likert – constat 1 (suite)

- Par exemple, une échelle comme celle-ci:

1. Super!	2. Hein? Je pense que je me suis perdu(e) en cours de route.	3. C'était dur, mais j'y suis arrivé(e).	4. Je ne pense pas que je vais y arriver.	5. Je sens que je vais être malade.

- Ressemblerait probablement plus à ceci:

1 2 3 4 5

Échelles de Likert – constat 1 (suite)

- Une recherche a montré que, par exemple, l'intervalle entre:

Jamais	Rarement	Souvent	Très souvent
1	2	3	4

1

- Ressemblait plutôt à:

Jamais	Rarement	Souvent	Très souvent

1,37

Bradburn et Sudman (1979)

Échelles de Likert – constat 1 (fin)

- Les scores globaux calculés selon la TCT, c'est-à-dire en faisant la somme des étiquettes numériques apposées aux catégories de réponses, donnent le même poids à chacune des catégories de réponses.
- Cependant, les catégories de réponses ne contribuent pas de la même façon à évaluer l'attitude des répondants.

Échelles de Likert – constat 2

- De la même façon, les items ne contribuent pas également à établir l'attitude des répondants.
- Être tout à fait d'accord avec:
  - Je fais de l'exercice régulièrement pour maintenir mon poids.
  - Je me fais vomir régulièrement pour maintenir mon poids.

N'entraîne pas le même diagnostic de santé...

Échelles de Likert –constat 3

- Les scores globaux de la TCT:
  - Permettent de faire une estimation du niveau « faible » ou « fort » d'une attitude.
  - Mais ils ne permettent pas de se faire une idée précise « de combien » plus fort est une attitude par rapport à une autre.

Échelles de Likert – autres constats

- De plus, ces scores:
  - ne possèdent pas la propriété d'addition;
  - ne permettent pas de faire des comparaisons facilement;
  - ne permettent pas de dresser un portrait détaillé de l'attitude des répondants;
  - de faire des prédictions.

Échelles de Likert – autres constats (fin)

- Enfin, ses scores sont dépendants:
  - Du groupe de répondants;
  - Des items du questionnaire.

Le modèle polytomique de Rasch pour les échelles de type Likert

Le modèle *Rating Scale*

Le modèle Rating Scale

- Permet d'étudier les items et la structure de l'échelle de réponses de façon plus approfondie.
- Le modèle considère les données brutes comme étant ordonnées (et non d'échelle intervalle ou ratio).
- Ainsi, il ne considère pas que l'intervalle entre chaque catégorie est égal; plutôt, il analyse les données et établit de quelle façon les catégories de l'échelle sont utilisées.

Le modèle Rating Scale (suite)

- Il transforme le nombre de fois qu'une catégorie de réponse est endossée en une échelle d'intervalle.
- Pour y arriver, le modèle introduit un nouveau paramètre:
  - Les seuils (« thresholds »).
- Il se sert du point où la probabilité d'opter pour la prochaine catégorie de réponse est égale à celle de conserver la précédente de telle sorte que les intervalles peuvent être interprétés comme le succès ou l'échec de passer à la catégorie suivante.

Le modèle Rating Scale (fin)

- L'échelle de réponses peut donc être représentée comme une succession de situations dichotomiques.
- Ainsi, en plus de fournir une estimation du niveau d'habileté des individus et du niveau de difficulté des items, le modèle présente une estimation du seuil entre chacune des catégories de l'échelle de réponses.

Exemple

Score F1 = 0 Score F2 = 0 Score F3 = 0 Score F4 = 0	Score F1 = 1 Score F2 = 0 Score F3 = 0 Score F4 = 0	Score F1 = 1 Score F2 = 1 Score F3 = 0 Score F4 = 0	Score F1 = 1 Score F2 = 1 Score F3 = 1 Score F4 = 0	Score F1 = 1 Score F2 = 1 Score F3 = 1 Score F4 = 1
F1	F2	F3	F4	

Rappel (modèle dichotomique)

Le modèle s'écrit:

$$Pni = \frac{e\,(Bn - Di)}{1 + e\,(Bn - Di)}$$

En prenant le logarithme naturel du rapport des chances, on obtient :

$$Ln\,[\,Pni / (1- Pni)\,] = Bn - Di$$

Modèle Rating Scale

Le modèle s'écrit:

$$Pni = \frac{e\,(Bn - Di - Fx)}{1 + e\,(Bn - Di - Fx)}$$

En prenant le logarithme naturel du rapport des chances, on obtient :

$$Ln\,[\,Pni / (1- Pni)\,] = Bn - Di - Fx$$

Applicabilité du modèle

- Modèle unidimensionnel.
- S'applique lorsqu'il y a plus de deux catégories de réponses.
- Et lorsque l'intervalle entre chaque catégorie de réponses de l'échelle demeure le même pour tous les items du questionnaire.

Hypothèses et propriété

- Comme pour le modèle dichotomique, l'unidimensionnalité et l'indépendance locale sont deux hypothèses fondamentales à respecter.
- De même, l'invariance peut être obtenue s'il y a adéquation entre les données et le modèle.
- Enfin, le modèle RS permet d'obtenir des estimations de paramètres séparées.

Ajustement des données au modèle

- Les mêmes indices d'ajustement que ceux utilisés pour le modèle dichotomique:
  - Indices infit et outfit (en version mnsq ou zstd) pour l'ajustement général, des items ou des personnes;
  - Erreur type moyenne;
  - Indice de fidélité;
  - Indice de séparation.

Analyses graphiques

- Il est également possible d'analyser le continuum graphiquement, de même que les catégories de réponses.

Critères d'analyse

- Tous les items servent à mesurer une même dimension.
- Chacune des catégories de réponse possède au moins dix observations.
- Les observations sont distribuées de façon uniforme dans les différentes catégories de réponses.

Critères d'analyse (suite)

- Indices infit et outfit dans l'intervalle (-2,2).
- Corrélation positive entre la mesure estimée par le modèle pour les personnes et la mesure observée. De même pour les items.
- Vérifier l'erreur type sur l'estimation, de même que l'erreur type moyenne.
- Indice de fidélité d'au moins 0,8. Indice de séparation plus grand que 2. (Vérifier le nombre de strata.)

Critères Likert (suite)

- La mesure moyenne estimée pour chacune des catégories devrait croître de façon monotone avec l'augmentation de la valeur de chacune des catégories de l'échelle de réponses. L'augmentation devrait être:
  - D'au moins 1 logit;
  - Mais de moins de 5 logits.

Critères Likert (suite)

- La cohérence mesure-catégorie et la cohérence catégorie-mesure.
- Vérifier les points d'ancrage de l'échelle de réponses.

Calibrage des points d'ancrage de l'échelle de réponses

- La valeur des points d'ancrage de l'échelle de réponses estimé par le modèle devrait augmenter en même temps que la valeur de chacune des catégories

Analyse graphique des catégories de réponses

- Permet de vérifier si la courbe de probabilités de chacune des réponses possède un sommet distinct et si l'ensemble des courbes apparait comme une suite de collines également distantes.
- Vérifier les points d'ancrage de l'échelle de réponses.

Le modèle *Rating Scale*

Un exemple

Données

- Questionnaire d'enquête auto-administré mis au point par le Centre de Formation Initiale des Maîtres (CFIM) de l'Université de Montréal
- Objectif: évaluer ses programmes de premier cycle universitaire en formation des maîtres
- Participants: Étudiant(e)s de 4e année du (B.EPEP).
- Projet de recherche subventionné par le Conseil de recherche en sciences humaines du Canada (CRSH) de 2002 à 2005 et de 2005 à 2008.

Questionnaire

- Première distribution du questionnaire: printemps 2000.
- Depuis, une opération de récolte de données a eu lieu chaque année, mais avec deux formes d'un même questionnaire (versions A et B).
- Initialement le questionnaire comportait huit sections. En 2004, il ne comportait plus que quatre sections:
  - perception générale de la formation
  - préparation à l'enseignement
  - les stages
  - et divers renseignements d'ordre démographiques

Items

- Dans la section du questionnaire retenue pour l'analyse de données, les étudiants doivent répondre à 20 items introduits par la phrase:
  - « Je considère que mon programme d'études m'a permis de développer des compétences pour… »
- Seule la version A des items est utilisée ici.

Échantillon d'items utilisés pour la présentation

2. Répondre aux questions des parents lors de la présentation du bulletin
3. Construire des outils pour l'évaluation sommative (contrôles, examens, etc.)
8. Planifier le déroulement d'activités d'apprentissage
10. Établir les règles de fonctionnement de la classe
12. Respecter les différences ethniques ou culturelles des élèves
19. Orienter les élèves vers les services d'aide appropriés
20. Discuter avec les parents des difficultés de leur enfant

Échelle de réponse

- En 2004, c'est une échelle de type Likert en six points qui a été utilisée.
- L'échelle est strictement positive et les catégories de réponses sont identifiées par des valeurs numériques.
- Seules la première et la dernière catégorie possèdent une étiquette.
- Ainsi, les catégories de réponses vont de «Tout à fait en désaccord = 1» à «Tout à fait en accord = 6».

Qualité de l'ajustement pour les personnes

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	person
1	22	7	-.56	.46	.64	-.6	.65	-.5	.96	28.6	41.9	1
2	33	7	1.70	.49	.34	-1.4	.40	-1.3	.94	57.1	50.0	3
3	20	7	.62	.45	.92	-1.5	.20	-1.6	.88	57.1	44.8	7
4	29	7	.82	.45	.40	-1.2	.34	-1.4	.83	71.4	48.2	8
5	28	7	.62	.45	4.43	3.7	4.82	3.8	-.10	28.6	44.8	9
6	25	7	.02	.44	.78	-.8	.86	-.1	.95	28.6	47.9	10
7	20	7	-.96	.46	.93	.1	.86	-.1	.93	42.9	41.7	12
8	16	7	-1.76	.47	.45	-1.1	.57	-1.0	.89	71.4	49.6	62
9	21	7	-.76	.44	1.35	.8	1.36	.8	.92	28.6	41.4	64
10	13	7	-2.63	.53	.60	-.8	1.03	.3	.64	42.9	57.1	67
11	32	7	1.47	.48	1.00	.0	1.14	.4	.80	57.1	52.4	69
12	22	7	-.56	.44	1.51	1.0	1.95	1.6	.46	14.3	41.5	70
13	18	7	-1.36	.45	1.16	.5	1.09	.4	.87	42.9	45.9	74
14	26	7	.22	.44	1.39	.9	1.19	.5	.81	42.9	45.8	76
15	15	7	-2.01	.48	.46	-1.0	.39	-1.2	.86	71.4	51.6	77
16	21	7	-.76	.44	.37	-1.4	.37	-1.4	.89	57.1	41.4	126
17	24	7	-.17	.44	1.27	.6	.99	.2	.73	57.1	48.3	146
18	36	7	2.50	.57	.53	-.7	.62	-.6	.59	57.1	54.3	147
19	19	7	-1.16	.45	.41	-1.2	.44	-1.2	.90	42.9	45.6	148
20	27	7	.41	.45	.49	-1.2	.56	-1.3	.87	71.4	47.8	150
21	10	7	-3.63	.70	3.07	2.1	1.69	.9	.45	71.4	72.1	160
22	17	7	-1.57	.46	.80	-.2	.85	-.1	.91	42.9	47.2	162
23	12	7	-2.83	.57	.21	-1.6	.25	-1.2	.90	85.7	62.5	164
MEAN	22.3	7.0	-.53	.48	.99	-.2	.97	-.2		50.9	48.6	
S.D.	6.8	.0	1.47	.06	.95	1.3	.94	1.2		17.8	7.3	

Qualité de l'ajustement pour les items

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	item
1	66	23	1.76	.29	.54	-1.4	.54	-1.4	.83	56.5	53.4	2. Répondre aux questions des pa
2	70	23	.16	.28	.96	-.4	.85	-.6	.75	39.1	46.6	3. Construire des outils pour l'é
3	105	23	-1.98	.27	1.49	1.4	1.73	2.0	.47	56.5	52.4	8. Planifier le déroulement d'ac
4	85	23	-.71	.28	.65	-1.1	.61	-1.4	.82	60.9	54.0	10. Établir les règles de fonctio
5	98	23	-1.50	.28	.95	-.3	.92	-.1	.76	47.8	53.8	12. Respecter les différences et
6	64	23	.51	.25	1.27	.9	1.06	.3	.74	43.5	46.9	19. Orienter les élèves vers les
7	66	23	1.76	.29	1.09	.4	1.09	.4	.71	52.2	53.4	20. Discuter avec les parents de
MEAN	79.4	23.0	.00	.28	.97	-.1	.97	-.1		50.8	48.6	
S.D.	21.8	.0	1.37	.02	.31	1.0	.37	1.1		7.2	3.9	

Sommaire pour les personnes

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	22.3	7.0	-.53	.48	.99	-.2	.97	-.2
S.D.	6.8	.0	1.47	.06	.95	1.3	.94	1.2
MAX.	36.0	7.0	2.50	.70	4.43	3.7	4.82	3.8
MIN.	10.0	7.0	-3.63	.44	.21	-1.6	.25	-1.6

REAL RMSE	.57	ADJ.SD	1.35
MODEL RMSE	.48	ADJ.SD	1.39
S.E. OF person MEAN = .31			

SEPARATION	2.39	person RELIABILITY	.85
SEPARATION	2.87	person RELIABILITY	.89

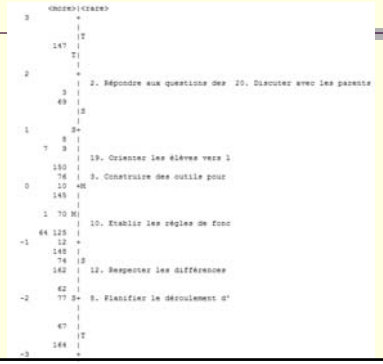
person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) person RAW SCORE RELIABILITY = .87

Sommaire pour les items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	73.4	23.0	.00	.26	.97	-.1	.97	-.1
S.D.	21.8	.0	1.37	.02	.31	1.0	.37	1.1
MAX.	105.0	23.0	1.76	.29	1.49	1.4	1.73	2.0
MIN.	46.0	23.0	-1.98	.24	.54	-1.6	.54	-1.4
REAL RMSE	.28	ADJ.SD	1.34	SEPARATION	4.82	item	RELIABILITY	.96
MODEL RMSE	.26	ADJ.SD	1.34	SEPARATION	5.13	item	RELIABILITY	.96
S.E. OF item MEAN = .86								

UMEAN=.000 USCALE=1.000
item RAW SCORE-TO-MEASURE CORRELATION = -1.00
161 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 350.06

Analyse du continuum



Analyse des points d'ancrage de l'échelle

SUMMARY OF CATEGORY STRUCTURE. Model="R"

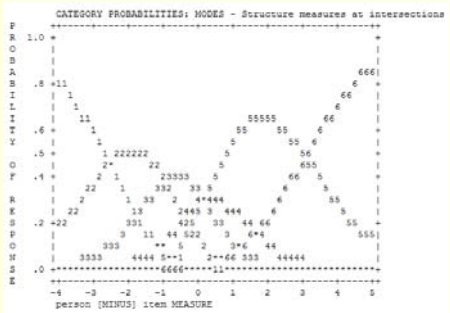
CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	%	SAMPLE AVRGE	EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE
1	1	29	18	-3.12	-3.00	.88	.89	NONE	(-3.77) 1
2	2	36	22	-1.63	-1.70	.90	.83	-2.53	-1.87 2
3	3	29	18	-.59	-.65	1.14	1.32	-.95	-.98 3
4	4	20	12	.21	.30	.75	.71	.19	.37 4
5	5	36	22	1.44	1.32	.81	.86	.20	1.82 5
6	6	11	7	2.22	2.53	1.64	1.20	3.09	(4.23) 6

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

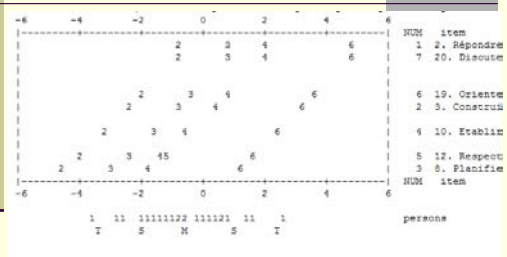
CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----	PROBABLTY	M->C	C->M	DISCR	
1	NONE		(-9.77)	-INF	-2.91		77%	58%	1
2	-2.53	.28	-1.87	-2.91	-1.15	-2.70	44%	47%	1.27 2
3	-.95	.25	-.58	-1.15	-.11	-1.06	40%	41%	1.07 3
4	.19	.26	.37	-.11	.95	-.04	33%	55%	.93 4
5	.20	.27	1.02	.93	3.23	.67	64%	61%	1.06 5
6	3.09	.38	(4.23)	3.23	+INF	3.14	75%	27%	.77 6

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

Analyse des points d'ancrage (suite)



Analyse de la structure des points d'ancrage



Conclusion

- Le modèle Rating Scale est utile pour l'analyse de questionnaires qui utilisent des échelles de type Likert.
- Il permet de mieux analyser les différents problèmes liés à ce type d'outil:
  - texte d'introduction
  - tâche à réaliser
  - libellé des énoncés
  - étiquette des points d'ancrage
  - et nombre de points.

Pour en savoir plus...

- Blais, J.-G., Grondin, J., Loye, N., Raîche G. (sous-presse). *Rasch model's contribution to the study of items and item response scales formulation in opinion/perception questionnaires*. Advances in Rasch Measurement: Volume Two.
- Blais, J.-G. et Grondin, J. (soumis). L'impact de la formulation des items dans les questionnaires d'enquête : une étude avec le modèle de Rasch pour les données polytomiques. *Mesure et évaluation en éducation.*
- Blais, J.-G. et Grondin, J. (soumis). The influence of labels associated with anchor points of Likert-type response scales in survey questionnaires. *Journal of Applied Measurement.*
- Grondin, J. et Blais, J.-G. (en préparation). How many response categories in Likert-type response scales used in survey questionnaires: a study with the Rasch model.

Le modèle *Partial Credit*

Très bref aperçu

Modèle Partial Credit

Le modèle s'écrit:

$$Pni = \frac{e\,(Bn - Di - Fix)}{1 + e\,(Bn - Di - Fix)}$$

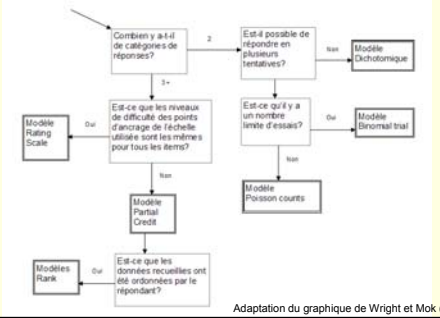
En prenant le logarithme naturel du rapport des chances, on obtient :

$$Ln\,[\,Pni / (1- Pni)\,] = Bn - Di - Fix$$

Survol des modèles unidimensionnels de Rasch

Les plus fréquents

Modèles unidimensionnels de Rasch les plus fréquents



Adaptation du graphique de Wright et Mok (2004), p. 16

Historiquement...

- Modèle dichotomique (Rasch, 1960)
- Approfondi (Wright, 1979)
- Étendu aux données de type Likert:
  - le modèle Rating Scale (Andrich, 1978)
- Modèle Partial Credit (Masters, 1982)
- Étendu aux situations de testing comportant plusieurs facettes:
  - le modèle Facets (Linacre, 1989)

Références

- Bradburn, N. M., & Sudman, S. (1979). *Improving interview method and questionnaire design : response effects to threatenning questions in survey research*. San Francisco: Jossey-Bass Publications.
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement: theory, models and applications* (pp. 258-278). Maple Grove, MN: JAM Press.
- Wright, B. D., & Mok, M. M. C. (2004). An overview of the family of Rasch measurement models. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement: theory, models and applications* (pp. 1-24). Maple Grove, MN: JAM Press.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model : fundamental measurement in the human sciences*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Merci!

julie_grondin@uqar.ca