

Partie 6

La validité

1. Définition

- La **validité** d'un instrument correspond au **degré auquel il mesure ce qu'il prétend mesurer.**
C'est le concept central de la psychométrie
- Comme la fidélité, la validité d'une mesure ne peut être qu'**estimée.**
Mais contrairement à la fidélité, elle ne peut pas être résumée par un seul indicateur : son appréciation se base sur plusieurs informations
- On distingue trois grandes approches de la validité :
 - ✓ la validité **critérielle**
 - ✓ la validité de **contenu**
 - ✓ la validité de **construit**

2. La validité critérielle

- **Première conception** historique de la validité
- La **validité critérielle** : degré auquel une mesure est liée avec une autre mesure indépendante et directe de la même caractéristique, appelée **critère**
« Un test est valide dès lors qu'il corrèle avec quelque chose. »
Guilford (1954)
- Ce type de validité est entièrement **empirique**
- On distingue deux types de validité critérielle :
 - ✓ la validité **concourante**
 - ✓ la validité **prédictive**

2. La validité critérielle

La validité concourante

- Le critère est un **état actuel** des sujets. C'est la capacité du test à détecter un état actuel des sujets : on parle également de **validité diagnostique**

Exemple :

On a construit un test mesurant les aptitudes en lecture.

Si ce test est valide, il doit détecter si un enfant appartient au groupe des enfants déficients en lecture ou au groupe des enfants normaux

2. La validité critérielle

- Lorsque le critère est **l'appartenance à un groupe**, on étudie la validité critérielle par la **méthode des groupes contrastés**

But : construire un outil valide en retenant les items qui différencient le plus un groupe représentatif d'une **population cible** d'un groupe représentatif de la **population normale**

- Le **MMPI** est l'inventaire le plus célèbre issu de cette méthode
 - ✓ MMPI = *Minnesota Multiphasic Personality Inventory*
 - ✓ inventaire de personnalité pathologique
 - ✓ créé en 1939 par Hathaway et McKinley
 - ✓ MMPI-2 (1989) : 567 items
(choix entre « Vrai », « Faux » ou « ne sait pas »)

2. La validité critérielle

- L'objectif était de construire un **outil diagnostic** qui :
 - ✓ ait la forme d'un **questionnaire d'auto-évaluation**
 - ✓ soit l'équivalent d'un **entretien psychiatrique structuré**
 - ✓ permette d'identifier les principales **pathologies mentales**
- Le MMPI a été développé suivant trois étapes :

Etape n°1 : la génération des items

Création d'un ensemble d'items sur la base de différents éléments : échelles existantes, rapports cliniques, observations de patients, ...

Génération d'environ 1000 items

ex : « Quand j'étais jeune, j'ai quelquefois volé des choses »

2. La validité critérielle

Etape n°2 : la création des groupes

Deux groupes ont ensuite été créés :

un **groupe clinique** (700 sujets pathologiques) : constitué de sous-groupes correspondant aux catégories nosographiques majeures connues à l'époque (dépression, hystérie, schizophrénie, ...)

un **groupe témoin** (sujets non pathologiques) : visiteurs et parents des patients, étudiants et personnel de l'université

2. La validité critérielle

Etape n°3 : la comparaison des groupes

Logique :

- ✓ on compare les réponses du groupe témoin avec celles d'un sous-groupe clinique à l'ensemble des items
- ✓ on retient les items qui **discriminent** le plus les deux groupes

ex : item « Quelqu'un m'en veut. »

taux de réponses « VRAI » :

- dans le groupe témoin : 24%
- dans le sous-groupe « Paranoïa » : 91%

2. La validité critérielle

- Validité critérielle du MMPI : méta-analyse de Hiller (1999)
 - ✓ MMPI (5007 protocoles) et Rorschach (2276 protocoles)
 - ✓ différents critères : diagnostic psychiatrique, auto et hétéro évaluation, comportement réel (ex : réponse au traitement)
 - ✓ résultats :
 - MMPI : 0.30**
 - Rorschach : 0.29**

2. La validité critique

Psychological Assessment
1999, Vol. 11, No. 3, 278–296

Copyright 1999 by the American Psychological Association, Inc.
1040-3590/99/\$3.00

A Comparative Meta-Analysis of Rorschach and MMPI Validity

Jordan B. Hiller
Harvard University

Robert Rosenthal
University of California, Riverside

Robert F. Bornstein
Gettysburg College

David T. R. Berry
University of Kentucky

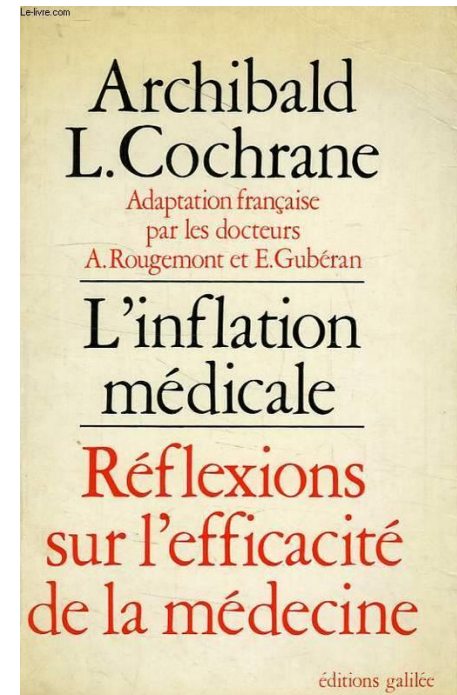
Sherrie Brunell-Neuleib
University of Oregon

Two previous meta-analyses concluded that average validity coefficients for the Rorschach and the MMPI have similar magnitudes (L. Atkinson, 1986; K. C. H. Parker, R. K. Hanson, & J. Hunsley, 1988), but methodological problems in both meta-analyses may have impeded acceptance of these results (H. N. Garb, C. M. Florio, & W. M. Grove, 1998). We conducted a new meta-analysis comparing criterion-related validity evidence for the Rorschach and the MMPI. The unweighted mean validity coefficients (\bar{r} s) were .30 for MMPI and .29 for Rorschach, and they were not reliably different ($p = .76$ under fixed-effects model, $p = .89$ under random-effects model). The MMPI had larger validity coefficients than the Rorschach for studies using psychiatric diagnoses and self-report measures as criterion variables, whereas the Rorschach had larger validity coefficients than the MMPI for studies using objective criterion variables.

La notion de méta-analyse

- Dans les années 1970-80, l'analyse gestionnaire de la Santé amène à questionner l'efficacité des pratiques médicales
 - ✓ les dépenses de santé sont-elles justifiées ?
 - ✓ la médecine soigne-t-elle réellement les gens ?

Contribution de l'épidémiologiste Cochrane :
Effectiveness and efficiency (1972)
(traduction française : *L'inflation médicale, réflexions sur l'efficacité de la médecine*)



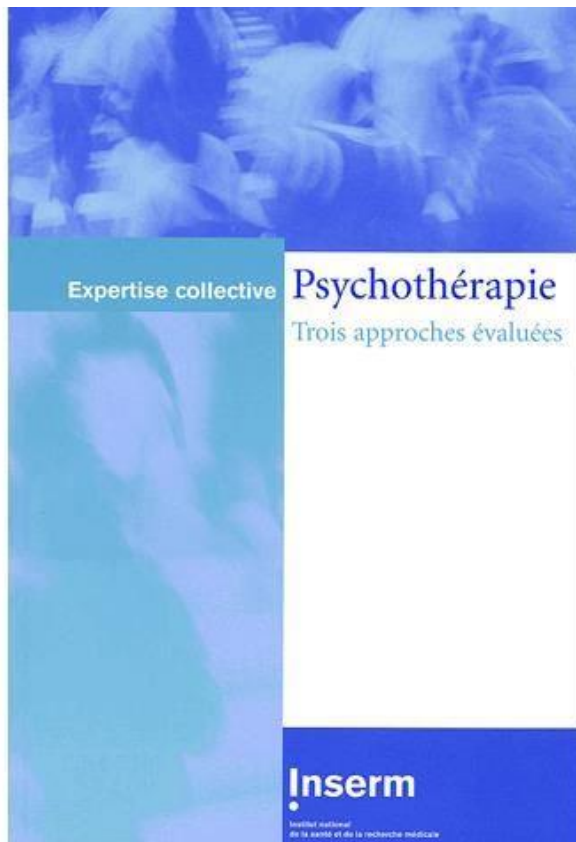
La notion de méta-analyse

- Dans ce contexte, les exigences concernant les **preuves médicales** deviennent plus fortes. On veut savoir si un traitement médical donné fonctionne ou pas.
- Mais la plupart du temps, un traitement médical a fait l'objet de plusieurs études dont les résultats sont parfois contradictoires. Comment statuer sur l'efficacité du traitement en question ?
- A la fin des années 1970, deux chercheurs, Gene Glass et Mary Lee Smith, ont proposé une méthode statistique apportant une solution à ce problème, la **méta-analyse**.

Réaliser la synthèse (quantitative) de l'ensemble des études portant sur le traitement afin de fournir une estimation numérique de son efficacité globale (la **taille d'effet**).

La notion de méta-analyse

- Les méta-analyses concernent l'efficacité des **outils diagnostics** et des l'efficacité des **traitements**



2. La validité critérielle

La validité prédictive

- Le critère est un **état futur** des sujets. C'est la capacité de la mesure à prédire un comportement futur des sujets : on parle également de **validité pronostique**

« Le coefficient de validité d'un test est le coefficient de corrélation entre la performance des sujets au test et la performance dans une activité que le test est censé prévoir. » Piéron (1979)

2. La validité critérielle

Exemple 1 : la validité prédictive du QI

Beaucoup d'études ont examiné dans quelle mesure le QI prédit différents critères liés à l'intelligence

- ✓ la **performance scolaire** : corrélation de 0.50
(le QI prédit 25% de la variance des résultats scolaires)
- ✓ le **revenu** : corrélation entre 0.40 et 0.50
(le QI prédit entre 16% et 25% de la variance du revenu)

2. La validité critérielle

Exemple 2 : la validité prédictive du « test du chamallow »

- Test comportemental créé en 1972 par Walter **Mischel**.
Un enfant a le choix entre manger un chamallow immédiatement ou attendre 15 min pour en avoir 2 (avec un chamallow sous les yeux)

2. La validité critérielle



2. La validité critique

Exemple 2 : la validité prédictive du « test du chamallow »

- 653 enfants de 4 ans avaient participé à l'expérience :
 - ✓ 1/3 des enfants avaient attendu quelques secs : les « **impulsifs** »
 - ✓ 1/3 des enfants avaient attendu quelques minutes
 - ✓ 1/3 des enfants avaient attendu les 15 min : les « **contrôlés** »

2. La validité critérielle

Exemple 2 : la validité prédictive du « test du chamallow »

- Ce test mesure la **capacité à différer la gratification** (attendre avant d'obtenir un plaisir, une récompense)
- Comparaison entre « impulsifs » et « contrôlés » 14 ans plus tard sur plusieurs critères, dont les scores au SAT.

Le **SAT** (*Scholastic Assessment Test*) est un test d'aptitudes scolaires américain composé de deux échelles, mathématiques et verbal (scores : de 600 à 2400)

- Résultats :
 - ✓ score moyen des **impulsifs** : **1052**
 - ✓ score moyen des **contrôlés** : **1262**

2. La validité critérielle

- On relève **3 difficultés** théoriques et méthodologiques auxquelles sont confrontées les études de validité critérielle :
 - ✓ la validité du critère
 - ✓ la contamination du critère
 - ✓ l'effet de sélection

2. La validité critérielle

La validité du critère

- Théoriquement, la validité critérielle n'est pas une démarche valable car elle entraîne un problème de **régression à l'infini**
 - ✓ on examine si une mesure corrèle avec un certain critère
Comment savoir si ce critère est valide ?
 - ✓ il faut regarder la corrélation de ce critère avec un autre critère
Comment savoir si ce dernier critère est valide ?
 - ✓ etc.
- Sur le plan pratique, on est obligé d'admettre qu'un certain critère est valide. En médecine, ce critère de référence est le **gold standard**

2. La validité critérielle

La contamination du critère

- La **contamination du critère** renvoie au cas où la mesure que l'on cherche à valider influence la mesure du critère.
Ce phénomène aboutit à une **augmentation artificielle** de la validité

Exemple :

Mesure de la satisfaction de vie

- ✓ t1 : rapporter des informations biographiques
- ✓ t2 : estimer son degré de satisfaction de vie

La deuxième mesure peut être influencée par la première.

2. La validité critérielle

L'effet de sélection

- Lorsque le critère est lié à un processus de sélection des sujets (scolarité, concours, recrutement,...), la validité critérielle est sous-estimée en raison d'une **restriction de la variance du prédicteur** : c'est l'**effet de sélection**

Exemple :

On cherche à prédire la moyenne en L2 à partir de la moyenne en L1

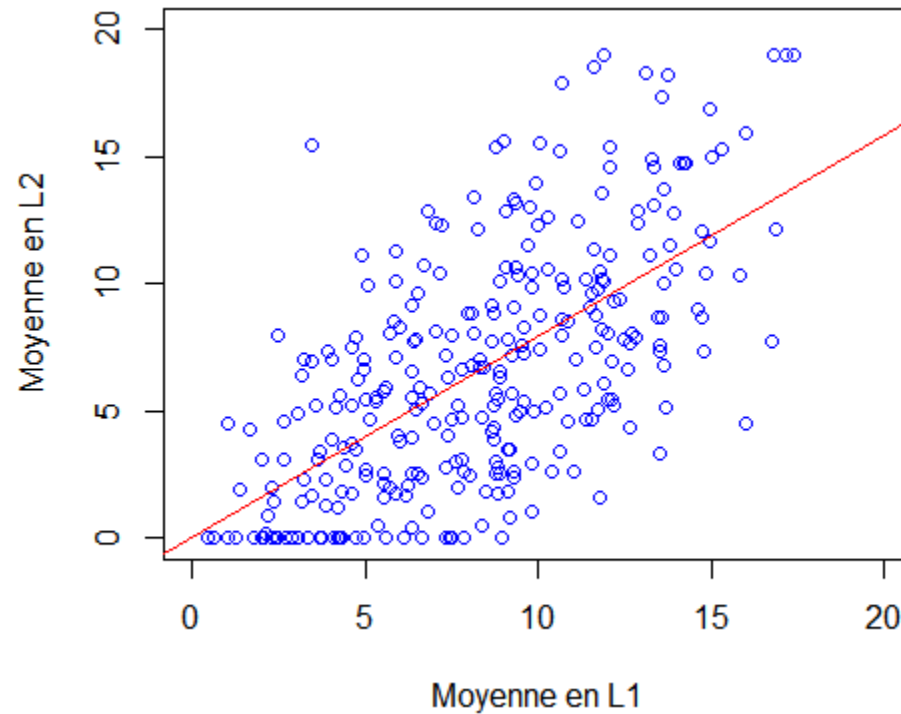
La corrélation ne peut être calculée que chez les L2

Cette corrélation est sous-estimée : la sélection entre la L1 et la L2 fait que chez les L2, la variance de la moyenne en L1 est réduite

2. La validité critique

L'effet de sélection

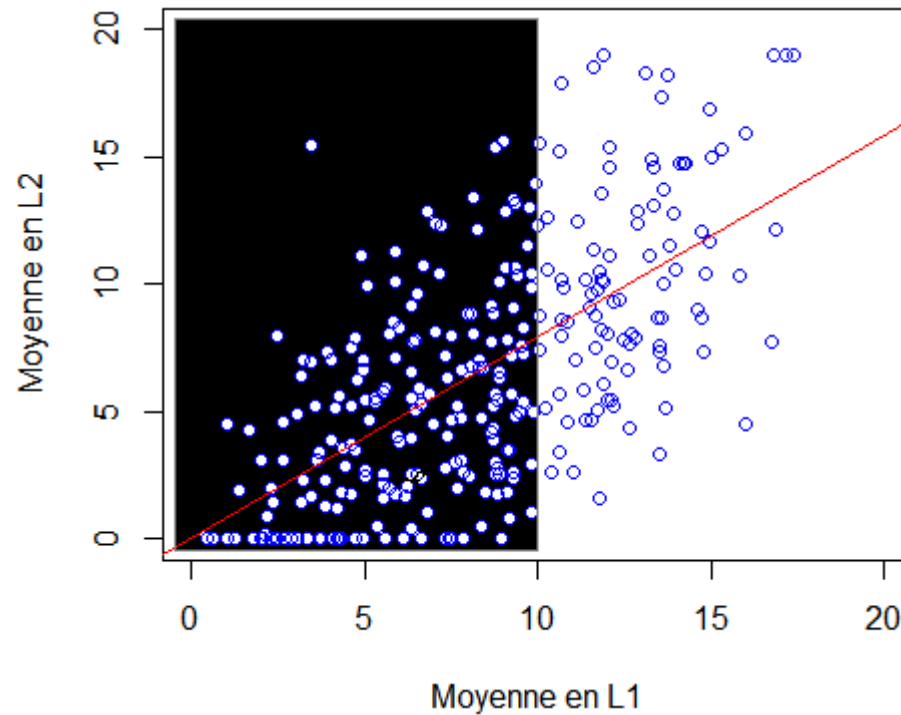
$$r = 0.64$$



2. La validité critique

L'effet de sélection

$$r = 0.35$$



3. La validité de contenu

- **Validité de contenu** : degré de représentativité des items en tant qu'échantillon du domaine de comportements à mesurer
- Un construit renvoie à un grand nombre d'observations et donc à un grand nombre d'items potentiels : **univers d'items**.
Les items retenus pour composer un instrument ne constituent qu'un **échantillon** de cet univers
- Un instrument est valide si ses items sont **représentatifs** de l'univers d'items relatif au construit qu'il mesure
- Deux méthodes principales pour l'étude de la validité de contenu :
 - ✓ l'**approche en facettes**
 - ✓ le **recours à des juges**

3. La validité de contenu

L'approche en facettes

- L'**approche en facettes** d'un construit consiste à le définir par les différentes sources de variation des réponses (**facettes**)
 - ✓ une définition d'un construit ayant ce format est appelée une **phrase en facettes**
 - ✓ un item = **combinaison** de certaines modalités des facettes
 - ✓ cette approche permet donc de **générer** l'univers d'items de façon exhaustive (ensemble de toutes les combinaisons)

3. La validité de contenu

Exemple : la mesure des valeurs

- Définition du construit par une phrase en facettes :

Une valeur est une conception individuelle :

- ✓ d'un **objectif (= facette 1)** : – terminal
– instrumental
- ✓ exprimant des **intérêts (= facette 2)** : – individuels
– collectifs
– mixtes
- ✓ relatifs à un **domaine** motivationnel (**= facette 3**) : – autonomie
– bienveillance
etc.

3. La validité de contenu

Exemple : la mesure des valeurs

- On peut ainsi générer l'univers des items (les différentes valeurs) en fonction de ces facettes.

Exemple :

		FACETTE		
		Objectif	Intérêt	Domaine
ITEMS	<i>Liberté</i>	terminal	individuel	autonomie
	<i>Amitié</i>	terminal	collectif	bienveillance

3. La validité de contenu

Le recours à des juges

- On demande à des juges d'évaluer le degré de représentativité de chaque item par rapport au construit
 - ✓ les juges utilisent un item de Likert
 - ✓ on calcule le degré de concordance entre les juges (**accord inter-juges**)
- Cette méthode requiert une définition théorique précise du construit (point de référence des juges)

4. La validité de construit

- Approché proposée par **Cronbach et Meehl (1955)**. L'idée est d'assurer **l'existence** du construit mesuré et éclairer sa signification
- Du concept de validité au concept de **validation** d'une mesure : processus d'accumulation d'éléments empiriques permettant de préciser la **signification** de la mesure

« la validation n'est rien d'autre, par essence, qu'une recherche scientifique de la signification des mesures » Messick (1989)

4. La validité de construit

- Dans cette perspective, l'approche centrale pour valider une mesure consiste à établir son **réseau nomologique**

La mesure acquiert progressivement sa validité si l'on peut construire **un réseau de relations avec d'autres mesures** (du même construit et d'autres construits)

4. La validité de construit

Exemple : validité de construit d'une mesure de l'anxiété

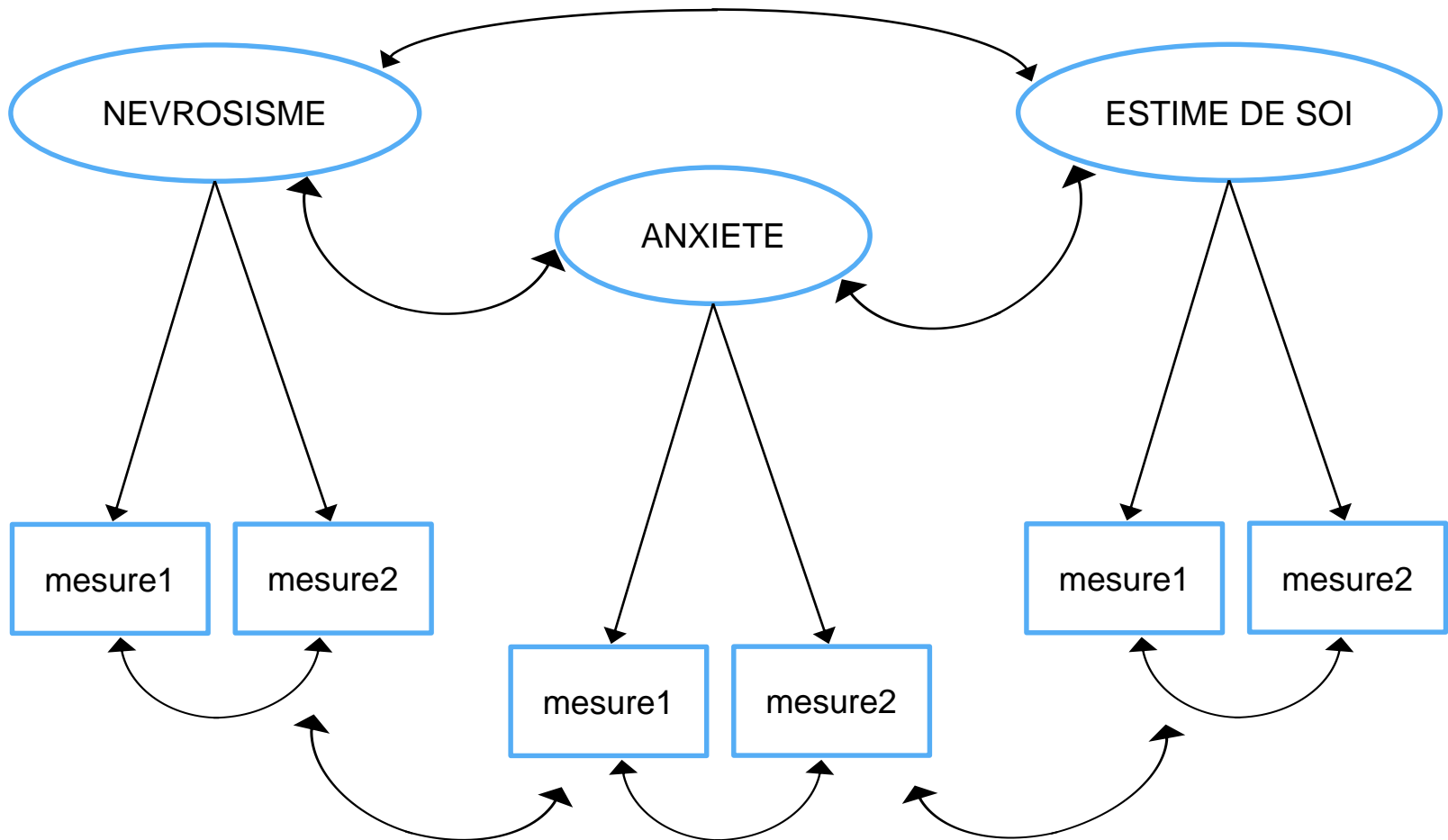
On s'attend théoriquement à ce que cette mesure :

- ✓ corrèle (positivement) avec des mesures du **névrosisme**
- ✓ corrèle (négativement) avec des mesures d'**estime de soi**
- ✓ ne corrèle pas avec des mesures d'**intelligence**
- ✓ soit liée de façon curvilinéaire avec des **performance scolaires**

C'est sur la vérification empirique de ces prédictions théoriques que s'appuiera la validité de la mesure développée

4. La validité de construit

Une partie du réseau nomologique de l'anxiété :



4. La validité de construit

- Deux types de validités liées à la notion de réseau nomologique :
 - ✓ la **validité convergente** : des mesures d'un même construit obtenues avec des méthodes différentes doivent corrélérer
 - ✓ la **validité divergente** : des mesures de construits différents ne doivent pas corrélérer
- Campbell et Fiske (1959) : la **matrice multitraits-multiméthodes**

4. La validité de construit

Exemple :

- ✓ trois construits : anxiété (A), dépression (D), névrosisme (N)
- ✓ deux méthodes : auto-évaluation (AE), hétéro-évaluation (HE)

		AE			HE		
		A	D	N	A	D	N
AE	A	0.9					
	D	0.12	0.92				
	N	0.1	0.82	0.95			
HE	A	0.72	0.06	0.03	0.91		
	D	0.01	0.7	0.62	0.11	0.92	
	N	-0.07	0.63	0.71	0.09	0.81	0.89

4. La validité de construit

Exemple :

✓ fidélité : corrélations **monotrait-monométhode**

		AE			HE		
		A	D	N	A	D	N
AE	A	0.9					
	D	0.12	0.92				
	N	0.1	0.82	0.95			
HE	A	0.72	0.06	0.03	0.91		
	D	0.01	0.7	0.62	0.11	0.92	
	N	-0.07	0.63	0.71	0.09	0.81	0.89

4. La validité de construit

Exemple :

✓ validité convergente : corrélations **monotrait-hétérométhodes**

		AE			HE		
		A	D	N	A	D	N
AE	A	0.9					
	D	0.12	0.92				
	N	0.1	0.82	0.95			
HE	A	0.72	0.06	0.03	0.91		
	D	0.01	0.7	0.62	0.11	0.92	
	N	-0.07	0.63	0.71	0.09	0.81	0.89

4. La validité de construit

Exemple :

✓ validité divergente (1) : corrélations **hétérotraits-monométhode**

		AE			HE		
		A	D	N	A	D	N
AE	A	0.9					
	D	0.12	0.92				
	N	0.1	0.82	0.95			
HE	A	0.72	0.06	0.03	0.91		
	D	0.01	0.7	0.62	0.11	0.92	
	N	-0.07	0.63	0.71	0.09	0.81	0.89

4. La validité de construit

Exemple :

✓ validité divergente (2) : corrélations **hétérotraits-hétérométhodes**

		AE			HE		
		A	D	N	A	D	N
AE	A	0.9					
	D	0.12	0.92				
	N	0.1	0.82	0.95			
HE	A	0.72	0.06	0.03	0.91		
	D	0.01	0.7	0.62	0.11	0.92	
	N	-0.07	0.63	0.71	0.09	0.81	0.89

Conclusion

- La validité n'est pas une propriété d'un test : elle renvoie davantage à la **pertinence des inférences** que l'on tire à partir des résultats au test, ou à l'interprétation des scores dans la cadre d'un **usage précis**

Conclusion

Exemple : validation d'un test d'arithmétique pour enfants

Usage du test : **prédire** la performance en mathématiques au collège

- ✓ interprétation : quelle sera la réussite future du sujet ?
- ✓ validité prédictive

Usage du test : **diagnostiquer** des difficultés d'apprentissage

- ✓ interprétation : quelles sont les difficultés spécifiques du sujet ?
- ✓ validité concourante

Usage du test : **caractériser** les stratégies de raisonnement

- ✓ interprétation : quels sont les processus cognitifs du sujet ?
- ✓ validité de construit

Conclusion

- La validité d'un test se rapporte non pas au test lui-même mais aux inférences faites à partir des résultats au test
- Il est inapproprié de parler de validité d'un test en général : un test peut être valide dans un contexte donné mais peu valide dans un autre contexte
- La validité n'est pas une affaire de « tout ou rien » : on parle de degré de validité
- Un test n'est jamais valide ou validé à 100%
- Un test acquiert sa validité par un cumul de preuves et non par le biais d'une seule étude