



Développement et validation de questionnaires et d'instruments

Atelier de formation A10

Plateforme de recherche clinique et évaluative

Anne-Sophie Julien, M.Sc., biostatistique



Objectifs d'apprentissage

1. Connaître les facteurs à considérer pour la création de questionnaire
2. Comprendre les étapes de développement d'un instrument psychométrique
3. Se familiariser avec les analyses statistiques permettant de valider un questionnaire



Plan de la formation

1. Développement de questionnaire
2. Développement d'instruments
3. Validation statistique



Questionnaire vs Instrument

Distinction pour le reste de la présentation:

- Questionnaire: Série de questions auxquelles on doit répondre, souvent à l'aide de choix de réponses
 - Exemples: Caractéristiques socio-démographiques, antécédents médicaux, habitudes de vie, etc.
- Instrument: Questionnaire visant à mesurer un concept non mesurable directement. Des valeurs numériques sont données aux réponses afin de quantifier le concept.
 - Exemples: Échelle de mesure sur la dépression, la qualité de vie ou l'anxiété



Section 1

Développement de questionnaire



Validité

Validité interne *(couverte dans la formation)*

- Le contenu du questionnaire est-il valide?

Validité externe *(non couverte dans la formation)*

- Est-ce que le questionnaire est complété par des participants de façon valide?
- Quelques éléments à considérer:
 - Inférence de l'échantillon à la population possible?
 - Répondants vs Non répondants, % de répondants
 - Caractéristiques des non répondants
 - Taille d'échantillon (test, retest, création / validation, études futures)
 - Présence de biais?



Validité interne

Les résultats obtenus sont-ils fiables?

- Absence de biais
- Cohérence et fiabilité entre les réponses (voir Section 3, plusieurs types de validité à vérifier)

Exemples de biais (il en existe un très grand nombre):

- Biais de rappel
 - Descriptions facilitantes
 - Ex.: questionnaire médication VIH au Brésil
- Biais de désirabilité sociale
 - Réponses mal perçues, ne pas oser répondre la vérité
 - Confidentialité à maximiser
 - Ex.: questionnaire consommation drogues



Qualités d'un questionnaire

Afin d'éviter les biais et augmenter la validité interne:

- Instructions claires
- Bonne longueur
- Questions regroupées par section (*NB: pas pour les instruments*)
- Formulation claire, neutre et précise des questions
- Choix de réponses adéquats
- Limiter la non-réponse
- Questions pour valider l'honnêteté des réponses et questions à réponses inversées



Types de questions à favoriser

1. Fermées

- Sélectionner un choix dans les catégories proposées
- Codification avec une variable qualitative

Exemples:

- Quel est votre groupe d'âge
(18-24) (25-44) (45-64) (65 et +)
- Avez-vous déjà reçu une injection de cortisone?
(Oui) (Non)



Types de questions à favoriser

2. À choix multiples

- Sélectionner un ou plusieurs choix parmi l'ensemble de catégories
- Créer une variable binaire pour chacun des choix possibles

Exemples:

- Quel(s) aliment(s) avez-vous consommé aujourd'hui?
(Bonbons) (Dessert) (Poulet pané) (Boisson sucrée) (Frites)
- Quel(s) médicament(s) avez-vous acheté dans la dernière année?
(Tylénol) (Advil) (Aspirin) (Voltaren) (Zantac) (Gravol)



Types de questions à favoriser

3. Échelles

- Likert
- Codification comme une variable ordinale (souvent distributions asymétriques, analyses données ordinales ou non paramétriques)
- Peut parfois être utilisée comme variable quantitative (distribution rarement normale)

Exemples:

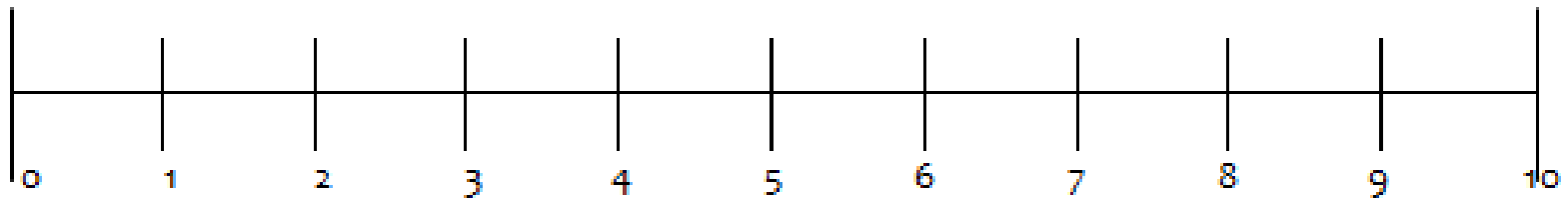
- Êtes-vous satisfait des soins reçus lors de votre dernière visite à l'urgence?
(1 Très insatisfait) (2) (3) (4) (5) (6) (7 Très satisfait)



Types de questions à favoriser

4. Échelle visuelle analogique

Sur une échelle de 1 à 10, comment considérez-vous la douleur actuelle?



Aucune
douleur

Pire
douleur
à vie

Encercler le chiffre ou faire un trait

Trait -> lecture avec règle ou lecteur numérique



Types de questions à éviter

1. Ouvertes

- Saisie complexe
- Regroupements à effectuer pour l'analyse quantitative
- Acceptable pour une analyse qualitative
- Acceptable si les réponses sont restreintes (ex.: âge)



Types de questions à éviter

1. Ouvertes

Exemple: De quelle maladie avez-vous déjà souffert?

- Répondant 1: Fibromialgie
- Répondant 2: Fibromyalgie
- Répondant 3: dépression, fibromyalgie
- Répondant 4: dépression post partum

Formulation préférable:

- Lister des groupes de maladies d'intérêt et demander si le patient en a déjà souffert



Types de questions à éviter

2. Classement

- Ordonner des items, préférences, etc.
- Analyse complexe
- Formulations alternatives



Types de questions à éviter

2. Classement

Exemple: Quelles sont dans l'ordre vos sources d'information médicale préférées?

(Info-Santé) (Médecin) (Pharmacien) (Livre) (Internet) (Entourage)

- Répondant 1: M, IS, P, E, L, I
- Répondant 2: I, E, L, IS, P, M

Formulation préférable:

- Quelle est votre source d'information préférée?
- À quelle fréquence utilisez-vous chacune de ces sources d'information?



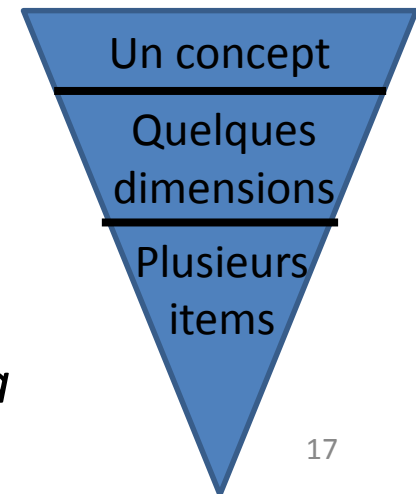
Section 2

Développement d'instruments



Instrument

- Outil, instrument de mesure, questionnaire psychométrique
- Permet de mesurer un concept abstrait (ou construit théorique) qui ne peut pas être mesuré directement
- Peut souvent être divisé en dimensions, sous-échelles ou construits
- Composé de plusieurs items, questions



Attention, le vocabulaire utilisé dans la littérature n'est pas toujours constant



Dimension

- Regroupement d'items permettant de mesurer un même aspect du concept
- Plusieurs questions (items) similaires afin de s'assurer que la dimension est bien mesurée

Item

- Question où les choix de réponses sont souvent sous forme d' « échelle de Likert »
- Peut aussi être un « Oui/Non » ou un « % d'accord »



Score

- Score = Calculé comme la somme des items ou leur moyenne, pour l'ensemble de l'instrument ou par dimension
- En présence de données manquantes, la somme est sous-estimée
 - Utiliser la moyenne calculée à partir du nombre d'items disponibles, si au moins 75% (règle du pouce) des items sont répondus (de l'instrument ou de la dimension)
 - Sinon, l'instrument (ou la dimension) est une valeur manquante



Étapes de développement

Plusieurs étapes plus ou moins facultatives qui dépendent du contexte:

- Création d'un nouvel instrument
- Validation d'un ancien instrument suite à des modifications
 - Traduction
 - Dans une nouvelle population
 - Dans un nouveau domaine



Étapes de développement

1. Déterminer les items pouvant être pertinents afin de mesurer le concept d'intérêt
 - Focus Group: Groupe de participants pour discuter du concept d'intérêt et faire une liste d'idées pertinentes (avoir a priori une idée des dimensions envisagées pour poser les bonnes questions)
 - Revue systématique: recherche d'instruments similaires dans la littérature afin d'en extraire leurs items
 - Possibilité de faire les deux, pour confirmer les résultats



Étapes de développement

2. Épurer les items

- De façon logique:
 - Éliminer les doublons, ceux trop différents, ceux qui ne semblent pas pertinents
 - Reformuler ceux qui ne sont pas clairs
- Collaborateurs
 - Avoir leur opinion pour avoir un tri plus précis
- Méthode Delphi
 - Classification des items (quantitative Likert 5 + commentaires) par des experts
 - Méthode itérative
 - Accepte ou rejette les items avec accord fort
 - Items avec accord mitigé: revus par tous avec les commentaires du 1er tour, et reclassifiés, jusqu'à convergence²²



Étapes de développement

3. Administration du questionnaire à un groupe test provenant de la population d'intérêt
4. Administration du questionnaire à nouveau sur le groupe test
 - Retest
5. Validation statistique du questionnaire
 - Plusieurs aspects à tester pour s'assurer de sa validité: voir la section 3
6. Peaufiner le questionnaire suite aux résultats
 - Supprimer ou reformuler des items



Exemple 1

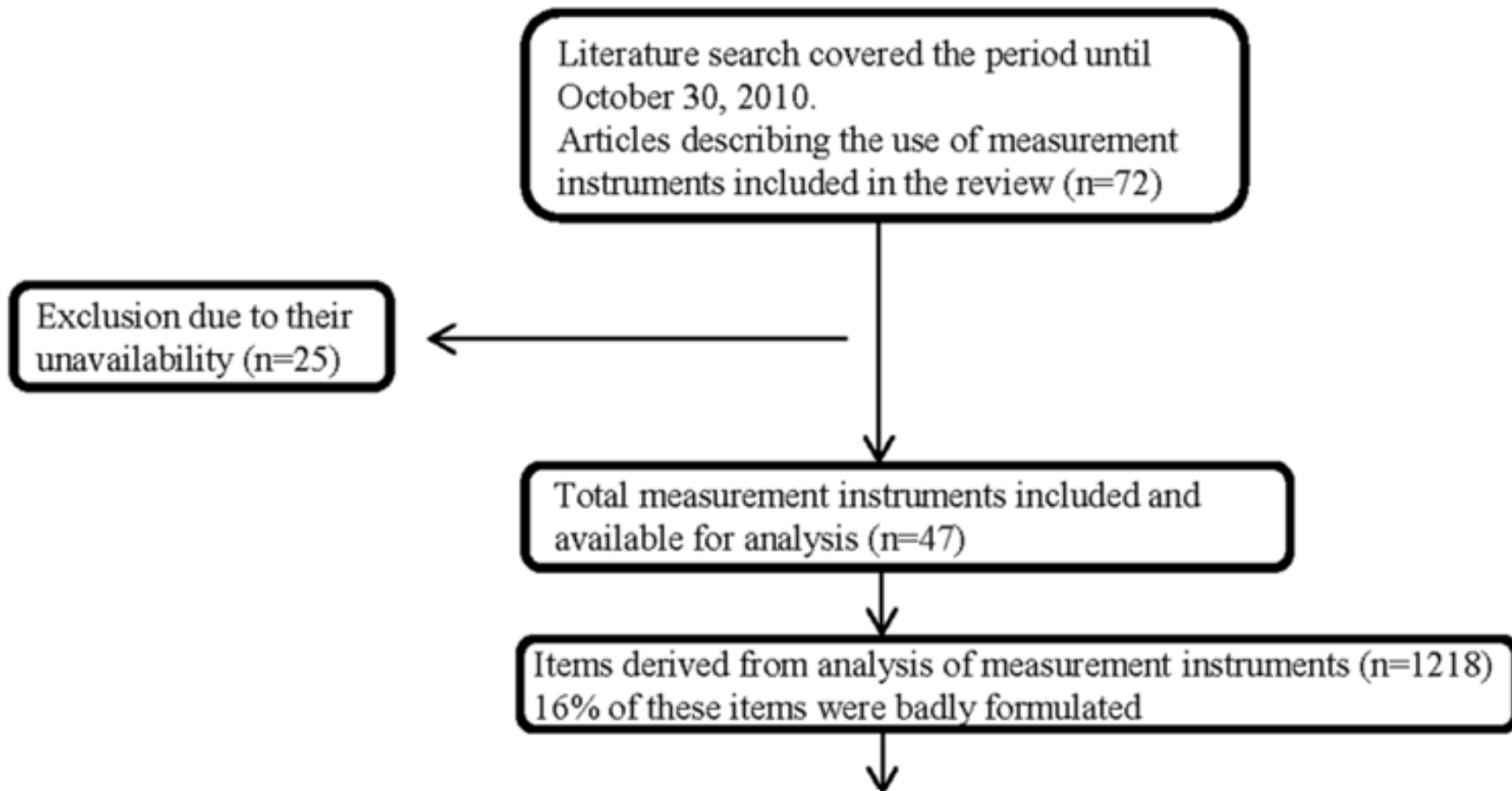
France Légaré, Francine Borduas, Adriana Freitas, André Jacques, Gaston Godin, Francesca Luconi, Jeremy Grimshaw, the CPD-KT team (2014). ***Development of a Simple 12-Item Theory-Based Instrument to Assess the Impact of Continuing Professional Development on Clinical Behavioral Intentions.*** PLOS ONE.

Contexte: Les décideurs dans les organisations fournissant du développement professionnel continu (DPC) ont identifié le besoin d'évaluer son impact sur la pratique. Le projet vise à développer un instrument global basé sur la théorie afin d'évaluer la répercussion des activités de DPC sur l'intention de comportement des professionnels de la santé.



Exemple

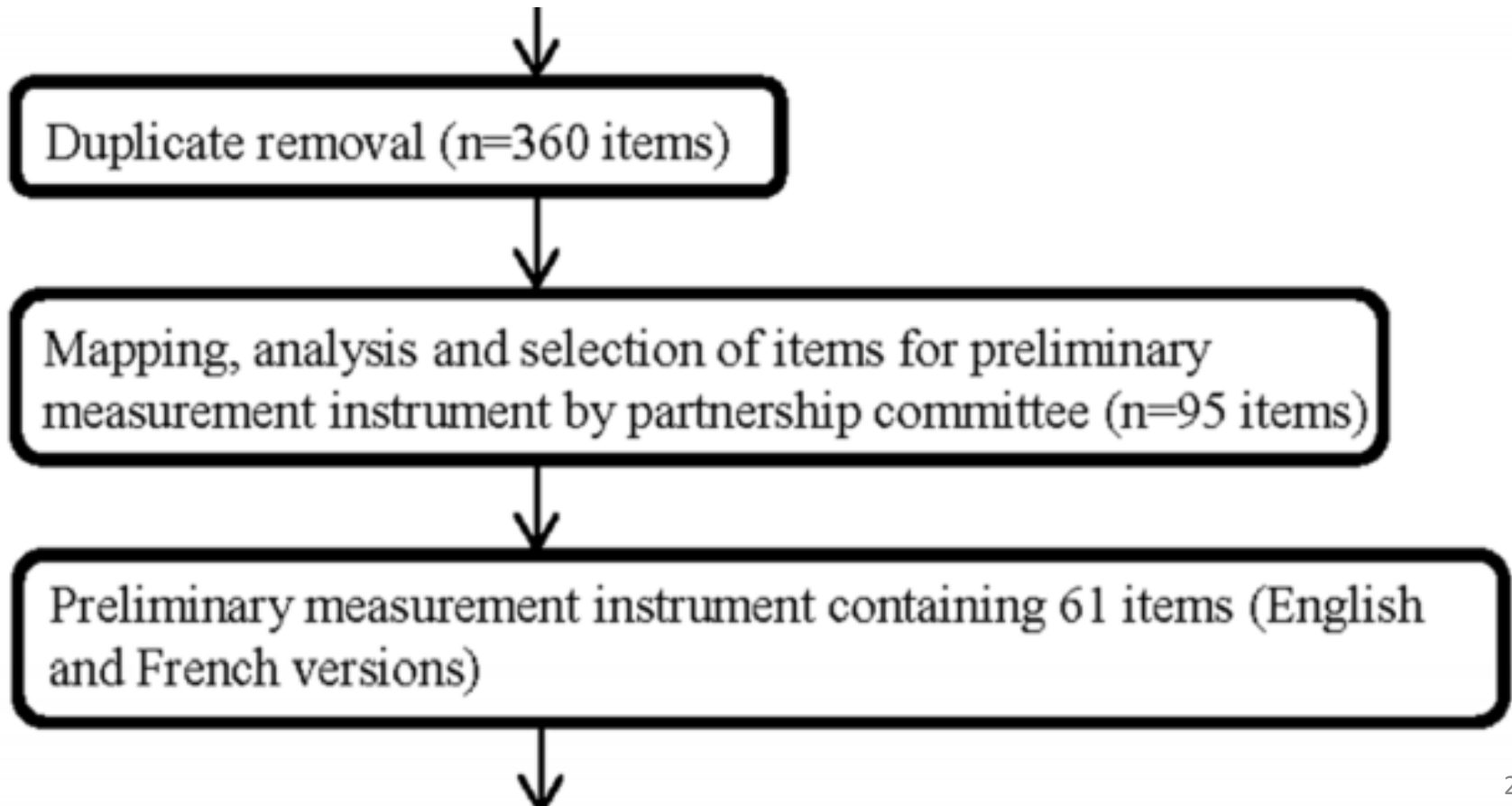
Étapes de développement: Phase 1





Exemple

Étapes de développement: Phase 2





Exemple

Étapes de développement: Phase 3

Synthesis of experts' responses in Round 1 (n=46 respondents)
Consensus achieved for 11 relevant items and 14 irrelevant items
36 items sent to experts

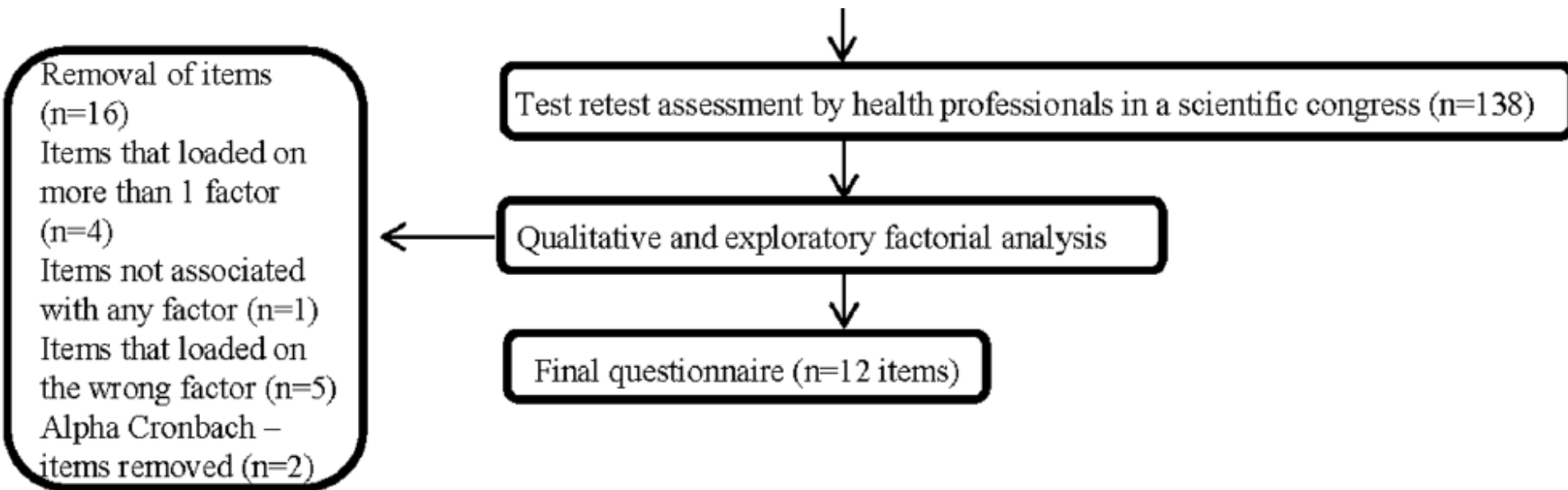
Synthesis of experts' responses in round 2 (n=41 respondents)
Consensus achieved for 15 relevant items and 10 irrelevant items
Partial consensus achieved on 11 items
Consensus achieved on the reformulation of 3 items for the
"moral norm" construct

Analysis of the e-Delphi results and provisional instrument proposed by
partnership committee resulted in 40 items



Exemple

Étapes de développement: Phase 4





Section 3

Validation statistique



Validation (1)

1. *Mettre les items dans le même sens avant l'analyse statistique*
2. Portrait des réponses
 - Est-ce que des variables devraient être recodées, y a-t-il des erreurs de saisie?
 - Tableaux de fréquences et statistiques descriptives
3. Honnêteté
 - Est-ce que le répondant a répondu de façon honnête et raisonnée aux questions?
 - Tableaux de fréquences croisés



Validation (2)

4. Cohérence interne (validity)

- Liens entre questions et dimensions
- Les items mesurent-ils une même dimension?
- Analyse factorielle exploratoire / confirmatoire
- Alpha de Cronbach

5. Stabilité temporelle, fiabilité (reliability)

- Mesure donnant le même résultat à plusieurs reprises
- Est-ce que les questions sont bien comprises?
- Test Retest avec des mesures d'accord / concordance
(Kappa, Intra-Classe, Bland-Altman)



Validation (3)

6. Validité convergente et divergente / discriminante

- Est-ce que les items de même dimension se ressemblent plus que des items de dimensions distinctes?
- Est-ce que les items de dimensions distinctes sont plus différents que des items de même dimension?

Corrélation



Validation(4)

7. Réactivité / “Responsiveness”

➤ Pour un concept qui évolue dans le temps, est-ce que l’instrument donne des résultats différents entre 2 mesures temporelles? (exemple: effet d’une intervention)

Taille d’effet

Test de différence de moyennes

8. Validité discriminante / “Known Group Validity”

➤ Est-ce que l’instrument donne des scores distincts à des groupes qui diffèrent selon le concept mesuré?

Test de différence de moyennes



Validation (5)

9. Validité concurrente / concomittante

➤ Est-ce que le nouvel instrument donne des scores semblables à ceux d'un instrument mesurant un concept équivalent?

Corrélation

Mesures d'accord / concordance

*** Attention, certains de ces termes sont utilisés de façon interchangeable dans la littérature***



Portrait des résultats

- Tableaux de fréquences pour chacune des variables du questionnaire
- Si très peu de variabilité, faire des regroupements (ex.: variable ordinale -> variable dichotomique)
- Si aucune variabilité, ou trop de valeurs manquantes, exclure la variable des analyses
- Bien connaître les données permet d'éviter des problèmes lors de la validation statistique



Honnêteté

- Est-ce que les participants ont répondu honnêtement, en prenant le temps de lire les questions avant de répondre?
- Croiser des variables qui sont reliées pour vérifier que les réponses font du sens
 - Variables qualitatives: Tableau de fréquences croisé
 - Variables quantitatives: Corrélations de Spearman / Pearson
- Exemple:
 - Déjà consommé de la drogue VS fréquence de consommation
 - Année de naissance vs année d'entrée à l'école
- Si on remet en doute les données d'un participant, le retirer pour la validation ou effectuer des analyses de sensibilité



ACP / AFE / AFC

- Création d'instruments -> Analyse en composantes principales
 - Pour déterminer le nombre de dimensions dans un instrument
 - Pour voir les items se regroupant dans une dimension
 - Étude pilote
- 1re validation d'instrument -> Analyse factorielle exploratoire
 - Pour vérifier que le nombre de dimensions dans l'instrument est identique aux versions précédentes
 - Pour s'assurer que les items contribuent aux dimensions attendues



ACP / AFE / AFC

- 2e validation d'instrument -> Analyse factorielle confirmatoire
 - Sur un autre échantillon que l'AFE (taille d'échantillon supérieure à l'AFE)
 - Pour vérifier que le modèle (dimensions + leurs items) déterminés dans l'AFE tiennent bien dans un autre échantillon de la nouvelle population
- L'ACP et l'AFE sont opposés d'un point de vue mathématique, mais d'un point de vue pratique, elles sont équivalentes



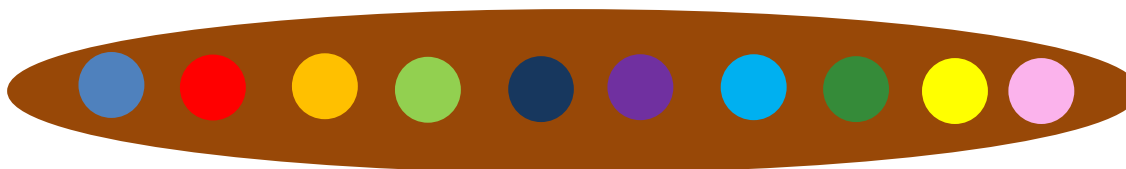
ACP / AFE

10 couleurs (items)

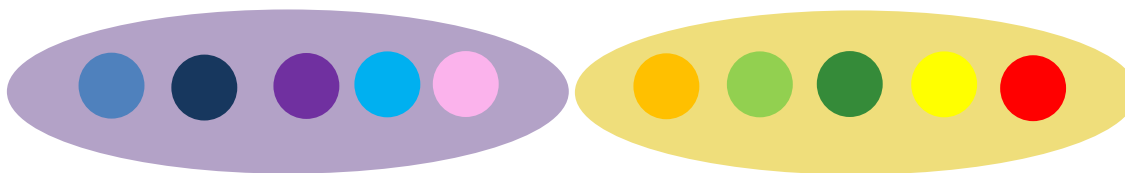


Combien de / Quels groupes (dimensions) peut-on faire sans trop de perte d'information?

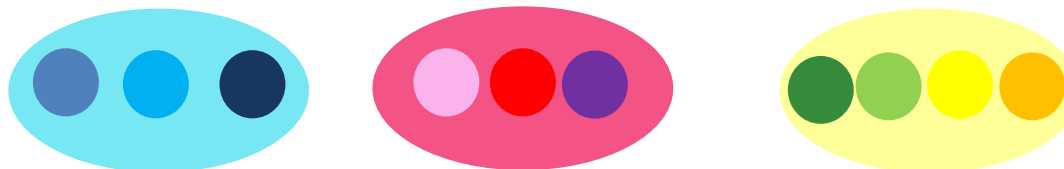
1 groupe: les couleurs



2 groupes:
Chaud vs Froid

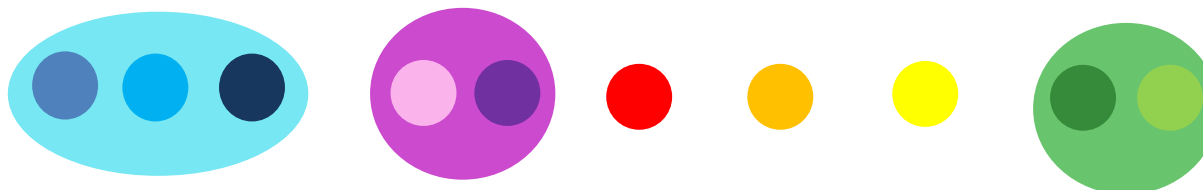


3 groupes:
Couleurs primaires



...

6 groupes:
Couleurs secondaires



...



ACP / AF

Choix du nombre de dimensions

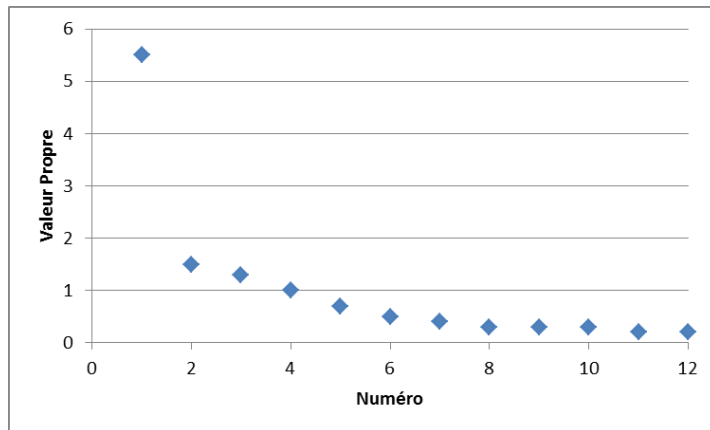
Plusieurs règles basées sur:

- Diagramme des valeurs propres (Catell / Éboullis / Scree Plot)
- Pourcentage de variation expliquée
- Valeurs propres
- Diffèrent selon les auteurs
- Arbitraires, subjectives
- Comparer le # proposé par les différentes règles
- Vérifier l'interprétation des dimensions pour aider au choix



ACP / AF

Choix du nombre de dimensions



- Diagramme de Catell:
avant le pied de l'éboulis
- Valeurs propres: > 1 ou > 0.7
- Proportion de variance:
 $> 10\%$ ou 15%
- Proportion de variance cumulée:
 $> 60\%$, 80% ou +

	Valeur propre	Différence	Proportion de variance	Proportion cumulée
1	5,5	4,0	45,5	45,5
2	1,5	0,2	12,4	57,9
3	1,3	0,3	10,4	68,3
4	1,0	0,3	8,1	76,4
5	0,7	0,2	5,4	81,8
6	0,5	0,1	4,0	85,8
7	0,4	0,0	3,2	89,0
8	0,3	0,0	2,8	91,8
9	0,3	0,1	2,8	94,6
10	0,3	0,1	2,3	96,9
11	0,2	0,0	1,7	98,6
12	0,2	-	1,4	100,0



ACP / AF

Interprétation de la solution factorielle

- « Loadings » = Corrélations entre l'item et la dimension
 - Près de 1 ou -1, item très important dans cette dimension
 - Près de 0, item non relié à cette dimension
 - >0.3 , 0.4 ou 0.5 , l'item "load" dans cette dimension (seuil diffère selon les auteurs)
- Valider les items contenus dans chaque dimension
- Avoir au moins 3 items par facteur
- Retirer les items qui se trouvent sur plus de 1 facteur
- Retirer les items qui se trouvent sur aucun facteur



ACP / AF

Interprétation des coefficients: Exemple

Item		Factor 1 (Beliefs about capabilities)	Factor 2 (Social influences)	Factor 3 (Beliefs about consequences)	Factor 4 (Moral norm)	Factor 5 (Intention)
Q7	I have the ability to [behavior] (strongly disagree/strongly agree)	0.74934	0.18021	-0.12599	0.21059	-0.05858
Q25	I am confident that I could [behavior] (strongly disagree/strongly agree)	0.78329	-0.09811	0.08429	0.03143	0.18647
Q31	For me, [behavior] would be (extremely difficult/extremely easy)	0.96592	-0.03217	0.04087	-0.17871	0.02051
Q5	To the best of my knowledge, the proportion of colleagues who will [behavior] would be: (0-100%)	-0.03024	0.81163	-0.04977	-0.14544	0.21287
Q8	Now think about a co-worker who you respect as a professional. In your opinion, does he/she [behavior] (never/always)	0.14637	0.87245	-0.04222	0.13894	-0.21688
Q10	Most persons who are important for me in the profession would [behavior] (strongly disagree/strongly agree)	-0.11913	0.74838	0.16943	-0.06198	0.22519
Q35c	Overall, I think that [behavior] is, for me: (useless/useful)	0.05487	-0.04288	0.94033	-0.01561	0.01417
Q35d	Overall, I think that is [behavior]: (harmful/beneficial)	-0.04608	0.06134	0.93328	0.05920	-0.04677
Q29	[Behavior] is the ethical thing to do (strongly disagree/strongly agree).	-0.07456	-0.10323	-0.08101	0.93819	0.17571
Q32	It would be acceptable to [behavior] (strongly disagree/strongly agree)	0.04682	0.05495	0.22567	0.77132	-0.03758
Q1	I intend to [behavior] (strongly disagree/strongly agree)	0.00976	0.12708	-0.09372	0.12053	0.83805
Q12	I plan to [behavior] (strongly disagree/strongly agree)	0.22201	0.01258	0.14262	0.07211	0.66377

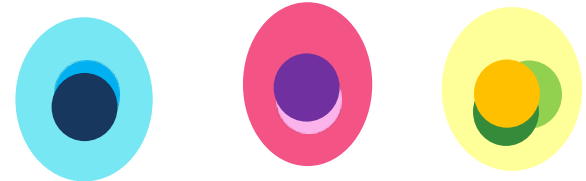


ACP / AF

Rotation

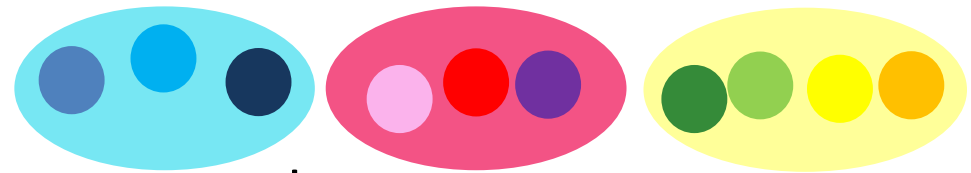
Avant rotation:

- Solution à 3 dimensions retenue
- L'interprétation n'est pas claire



Après rotation:

- Transformation linéaire
 - Regarder les données sous un autre angle
 - Pour simplifier l'interprétation
 - Pour mieux exploiter la solution
-
- Plusieurs types de rotations orthogonales (facteurs indépendants, ex.: Varimax) ou obliques (facteurs dépendants, ex.: Oblimin)
 - Nouveaux coefficients et nouvelles variances





AFC

- Valider si la solution factorielle (les dimensions choisies et les items se trouvant dans chacune) retenue par l'AFE tient la route dans un nouvel échantillon
- Déterminer si l'ajustement du modèle est bon à l'aide de différents critères

Critère	Bonnes valeurs
CFI (Comparative Fit Index)	> 0.94
SRMR (Standardized Root Mean Square Residual)	<0.055
RMSEA (Root Mean Square Error of Approximation)	<0.055 et BS IC à 90% <0.09
NNI (Bentler & Bonett's Non-normed Index)	>0.90
NFI (Bentler & Bonett's Fit Index)	>0.90
GFI (Goodness of Fit Index)	>0.90
Statistiques T des paramètres	>1.96 ou <-1.96



Alpha de Cronbach

- Cohérence interne entre les items d'un concept
- Fonction du nombre d'items et de leur variance individuelle et commune
- Faible coefficient: un item ne mesure pas le même concept que les autres
- Coefficient de 0 à 1
- Règle du pouce: Vise au moins 0.7, idéalement entre 0.8 et 0.9



Alpha de Cronbach

Variables	Alpha	Variable supprimée	Variables brutes		Variables centrées-réduites	
			Corrélation avec total	Alpha	Corrélation avec total	Alpha
Brut	0,78					
Normalisé	0,79	Q5	0,63	0,71	0,63	0,72
		Q8	0,64	0,70	0,65	0,70
		Q10	0,62	0,70	0,62	0,73

Facteur 2: Social Influences

Q5: To the best of my knowledge, the proportion of colleagues who will apply the objectives of the activity would be __%

Q8: Now think about a co-worker who you respect as a professional. In your opinion, would he apply the objectives of the activity?

Q10: Most persons who are important for me in the profession would apply the objectives of the activity



Test-Retest

- Faire remplir un questionnaire aux mêmes participants à deux reprises, après un certain laps de temps.
- Pour les questions intemporelles, on devrait s'attendre à ce qu'une personne réponde de la même façon aux deux temps.
- Si la cohérence entre les deux temps n'est pas bonne, cela pourrait montrer qu'une question est mal posée et ou mal comprise.
- Analyse grâce aux mesures d'accord / concordance:
 - Variables qualitatives -> Coefficient Kappa
 - Variables quantitatives - > Coefficient intra-classe



Coefficient Kappa de Cohen

- Calculé à partir d'un tableau croisé des deux variables qualitatives ayant les mêmes catégories (mesurées à 2 reprises)
- Est-ce que les deux mesures sont en accord?
- Calcul le nombre d'observations qui se trouvent sur la diagonale, soient les paires qui concordent
- Le kappa pondéré tient compte des accords parfaits, mais aussi des accords imparfaits en pondérant selon la distance entre les discordances



Kappa (Exemple)

		Satisfaction au retest					Total
		0	1	2	3	4	
Satisfaction au test	0	25	5	3	1	0	34
	1	3	20	4	2	1	30
	2	2	2	15	3	4	26
	3	3	1	3	10	2	19
	4	0	2	1	1	5	9
Total		33	30	26	17	12	118

Valeur	Interprétation
<0.2	Médiocre
[0.2, 0.4[Faible
[0.4, 0.6[Modéré
[0.6, 0.8[Bon
[0.8, 1.0]	Excellent

Kappa = 0,53
 Kappa pondéré = 0,59



Kappa (Exemple 1)

Variables = Likert 1 à 7 -> Coefficient Kappa pondéré

Valeurs modérées (0.40 à 0.60, aucun item retiré)

Item	Test			Retest			Reliability (weighted kappa)
	N	Mean (sd)	Median (min-max)	N	Mean (sd)	Median (min-max)	
Q1	137	6.14 (1.28)	7 (1-7)	121	6.02 (1.27)	6 (2-7)	0.54
Q2	137	6.26 (1.28)	7 (1-7)	121	6.17 (1.16)	7 (2-7)	0.56
Q4	137	5.55 (1.55)	6 (1-7)	121	5.55 (1.43)	6 (2-7)	0.51
Q5	136	4.23 (1.54)	5 (1-6.5)	121	4.22 (1.52)	5 (1-6.5)	0.45
Q7	138	5.38 (1.22)	5 (1-7)	121	5.60 (1.14)	6 (2-7)	0.53
Q8	138	5.54 (1.12)	6 (2-7)	120	5.57 (1.17)	6 (2-7)	0.41
Q10	137	5.13 (1.38)	5 (1-7)	121	5.27 (1.18)	5 (2-7)	0.42
Q12	137	6.14 (1.18)	7 (1-7)	121	5.99 (1.18)	6 (2-7)	0.60
Q13	138	6.35 (0.97)	7 (1-7)	121	6.32 (0.93)	7 (2-7)	0.52
Q15	138	3.43 (1.87)	3 (1-7)	121	3.26 (1.79)	3 (1-7)	0.41
Q16	138	6.21 (1.14)	7 (1-7)	121	6.13 (1.21)	7 (1-7)	0.59
Q18	138	6.01 (1.22)	6 (1-7)	121	6.03 (1.10)	6 (1-7)	0.50
Q21	137	5.32 (1.61)	6 (1-7)	120	5.14 (1.58)	5 (1-7)	0.50



Coefficient de corrélation intra-classe

- Intraclass Correlation Coefficient (ICC)
- Pour quantifier l'accord ou la concordance entre plusieurs évaluations par une variable continue
- Coefficient de 0 à 1
- Près de 1: grande consistance
- Plusieurs versions
 - Toujours spécifier laquelle est utilisée
 - Pas de grille d'interprétation
 - Dépendent de 3 paramètres (voir diapositive suivante)



Coefficient de corrélation intra-classe

Plusieurs définitions, selon:

1. Choix des juges
 - Échantillon parmi ceux disponibles ou les seuls d'intérêt
 - Les mêmes pour toutes les évaluations ou non

2. Mesure individuelle ou moyenne
 - On s'intéresse aux valeurs individuelles ou de façon générale

3. Accord ou consistance
 - Réponse exactement identique ou dans le même ordre d'idée

➤ Consulter un statisticien pour plus d'information sur ce calcul



Exemple 2

Bairati et al.: Development and validation of an instrument assessing women's satisfaction with screening mammography in an organized breast cancer screening program. BMC Health Services Research 2014 14:9. doi: 10.1186/1472-6963-14-9.

Objectif: Développer et valider un instrument de mesure en français pour évaluer la satisfaction des femmes ayant une mammographie de dépistage.



Validité convergente et discriminante

- Paires convergentes: items avec le reste de sa dimension
- Paires discriminantes: items avec les autres dimensions
- Présenter les étendues (minimum, maximum) des corrélations des paires convergentes et discriminantes
- Scaling success: Compare chaque paire de corrélations. Succès si la corrélation convergente > corr. discriminante

Dimensions	No. of items	Convergent validity (range of correlations)	Discriminant validity (range of correlations)	Scaling Success n/N ¹ (%)
Staff's communication skills	3	0.75-0.80	0.58-0.69	9/9 (100%)
Information given by the program	3	0.75-0.78	0.59-0.67	9/9 (100%)
Technician's skills	4	0.75-0.85	0.49-0.67	12/12 (100%)
Physical environment	4	0.71-0.74	0.52-0.66	12/12 (100%)

¹Number of convergent correlations significantly higher than discriminant correlations/total number of correlations.



Réactivité

- Pour des concepts qui changent dans le temps, il devrait y avoir une différence temporelle si le questionnaire est administré à 2 reprises
- Pour quantifier la réactivité:
 - Taille d'effet = $(\text{Moyenne après} - \text{Moyenne avant}) / \text{Écart-type}$
- Pour tester la différence de moyennes entre les deux temps
 - Test T Paired ou Rangs Signés de Wilcoxon
 - * Voir l'atelier 02 pour plus de détails **
- Une grande taille d'effet ou une différence significative entre les deux mesures indiquent que l'instrument répond bien au changement



Validité “Known-Group Validity”

- Vérifier si le score moyen de l’instrument diffère entre des groupes distincts
- Exemple: les scores à un instrument sur la dépression devrait être plus élevés chez des patients pour une dépression majeure par rapport à des patientes pour un accouchement
- Avec un score continu, deux groupes à comparer
 - Test T ou Wilcoxon
- Si plus de deux groupes à comparer
 - Analyse de la variance (ANOVA) ou Kruskal-Wallis

** Voir l’atelier 02 pour plus de détails **



Validité concurrente

- Si d'autres instruments reconnus mesurent des concepts semblables, ils devraient être corrélés avec l'instrument à tester
- Pour examiner la corrélation entre deux instruments
 - Coefficient de Corrélation de Pearson / Spearman
- Pour voir si deux instruments ont été répondus de la même façon (si leur échelle est la même)
 - Mesures d'accord: Kappa ou Coefficient de corrélation intra-classe



Conclusion

- Importance d'un choix réfléchi de questions et réponses
- Développement d'instrument plus complexe qu'un "simple" questionnaire
- Plusieurs étapes de validation interne
- Validation externe, un autre aspect important (non traité ici)



Références

- Légaré F. et al (2014). Development of a Simple 12-Item Theory-Based Instrument to Assess the Impact of Continuing Professional Development on Clinical Behavioral Intentions. PLOS ONE.
- Bairati I. et al.: Development and validation of an instrument assessing women's satisfaction with screening mammography in an organized breast cancer screening program. BMC Health Services Research 2014 14:9. doi: 10.1186/1472-6963-14-9.
- Fayers, P.M. , D. Machin (2000). Quality of Life: Assessment, Analysis and Interpretation. Chichester, England: John Wiley & Sons Ltd.
- Kline, P. (2000). The handbook of psychological testing (2nd ed.). London: Routledge, page 13
- Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". *Biometrics* 33 (1): 159-174. doi:10.2307/2529310.
- O'Rourke Norm and Larry Hatcher (2013). A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling, 2nd Edition. Cary, NC: SAS Institute Inc.
- STATISTIQUE CANADA. 2003. *Méthodes et pratiques d'enquête*, produit N°12-587-X au catalogue de Statistique Canada, Ottawa, Ontario, <http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-fra.pdf> (site consulté le 16 mars 2015).



Merci pour votre attention!

Des questions?