



Αξιοπιστία – Εγκυρότητα της Μέτρησης

**Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης
Δημοκρίτειο Πανεπιστήμιο Θράκης
Αλεξανδρούπολη
2012 - 2013**

Η παρουσίαση βασίζεται σε υλικό από...

- Cohen, L., Manion, L. & Keith, M. (2008)

Μεθοδολογία Εκπαιδευτικής Έρευνας, Αθήνα:
Μεταίχμιο

- <http://www.psych.uoa.gr/~roussosp/stats>



Ψυχομετρικά Χαρακτηριστικά της Μέτρησης

Αξιοπιστία

Εγκυρότητα

Βασικές Μέθοδοι Συλλογής Δεδομένων

- **Ερευνητικός σχεδιασμός:** το σχέδιο που ακολουθεί ο ερευνητής για να οργανώσει και να διεκπεραιώσει τη συλλογή των δεδομένων της έρευνας.

Βασικές ερευνητικές προσεγγίσεις συλλογής των εμπειρικών δεδομένων (με σκοπό είτε την περιγραφή νέων δεδομένων είτε τον έλεγχο υποθέσεων):

- Περιγραφική προσέγγιση (observational approach)
- Συσχετιστική προσέγγιση (correlational approach)
- Προπειραματική προσέγγιση (pre-experimental approach)
- Ψευδο-πειραματικοί σχεδιασμοί (quasi-experimental approach)
- Πειραματικοί σχεδιασμοί (experimental approach)

Ανεξάρτητα από το ποια μέθοδος θα χρησιμοποιηθεί, ο ερευνητής θα πρέπει να προσπαθεί να ελαχιστοποιήσει το σφάλμα της μέτρησης με το να ελαχιστοποιεί, όσο το δυνατόν, την πιθανότητα επίδρασης μη ελεγχόμενων παραγόντων στα αποτελέσματα.

Σφάλμα Μέτρησης (Measurement Error)

- **Σφάλμα μέτρησης:** η διαφορά της παρατηρούμενης τιμής από την πραγματική τιμή.

Δύο είδη σφαλμάτων:

- **συστηματικό σφάλμα:** όταν οι τιμές μεταβάλλονται με σταθερό (συστηματικό) τρόπο, δηλαδή οι παράγοντες που προκαλούν το σφάλμα επιδρούν με μια σταθερή τάση (π.χ. λάθος οδηγίες συμπλήρωσης, όταν οι συμμετέχοντες δίνουν κοινωνικά αποδεκτές απαντήσεις)
- **τυχαίο σφάλμα:** όταν οι τιμές μεταβάλλονται τυχαία, δηλαδή δεν υπάρχει κάποια συγκεκριμένη τάση στην επίδραση των παραγόντων που προκαλούν το σφάλμα. Συνήθως έχει να κάνει με απρόσμενες μεταβολές στις περιστάσεις κάτω από τις οποίες γίνονται οι μετρήσεις (πχ. αλλαγές στη διάθεση των συμμετεχόντων, αυξομείωση της προσοχής, αλλαγές στη θερμοκρασία, το φωτισμό, θόρυβος κλπ).

Τεχνικές Μείωσης Σφάλματος Μέτρησης

- Κατάλληλη εκπαίδευση των ερευνητών, προσφορά κινήτρων
- Έλεγχος των συνθηκών διεξαγωγής της μέτρησης από τους ερευνητές
- Διεξαγωγή πιλοτικής έρευνας για τη δοκιμή του εργαλείου μέτρησης σε δείγμα του υπό μελέτη πληθυσμού προκειμένου να εντοπισθούν δοκιμασίες/ερωτήματα που περιέχουν ασάφειες, προκαλούν δυσφορία, κόπωση, για κάποιο λόγο αυξάνουν το σφάλμα της μέτρησης).
- Έλεγχοι αξιοπιστίας και εγκυρότητας των ψυχομετρικών εργαλείων

Ψυχομετρικές Κλίμακες Μέτρησης

- Τα ψυχομετρικά εργαλεία είναι ερωτηματολόγια που κατασκευάζονται με βασικό σκοπό να προσεγγίσουν, δηλαδή να ποσοτικοποιήσουν, διάφορες όψεις του ανθρώπινου ψυχισμού.
- Ο αποκρινόμενος καλείται να συγκρίνει την συμπεριφορά ή τις συνήθειές του με αυτές που περιγράφει το εκάστοτε ερώτημα και να απαντήσει σε μία δίτιμη κλίμακα (θετικά-αρνητικά) ή σε κλίμακες τύπου Likert τη συχνότητα με την οποία ακολουθεί την περιγραφόμενη συμπεριφορά.
- Σε αρκετές περιπτώσεις ένα ερωτηματολόγιο είναι πολυδιάστατο, αποτελούμενο από ερωτήματα που αναφέρονται σε διαφορετικές συμπεριφορές ή χαρακτηριστικά και ονομάζονται υποκλίμακες ή υποδοκιμασίες. Δηλαδή, τα ερωτήματα αθροίζονται σε μικρότερες ομάδες που κάθε μία μετρά ένα χαρακτηριστικό της ανθρώπινης συμπεριφοράς.

Ψυχομετρικές Κλίμακες Μέτρησης

- Είναι γνωστό πως ο ανθρώπινος ψυχισμός αλλά και η συμπεριφορά επηρεάζονται από το στενό και ευρύτερο περιβάλλον. Έτσι, καθώς οι κοινωνικό-πολιτισμικές και οικονομικές συνθήκες αλλά και η κουλτούρα διαφέρουν ανάμεσα στα υποκείμενα, είναι εύλογο πως η χρήση ενός ερωτηματολογίου σε δείγμα ενός πληθυσμού με διαφορετικά χαρακτηριστικά από αυτόν για τον οποίο (και επί του οποίου) δημιουργήθηκε, προϋποθέτει τον επανέλεγχο καλής εφαρμογής του ή και την τροποποίησή του.
- Αυτό επιτυγχάνεται με την βοήθεια **της ψυχομετρικής θεωρίας**, όπου ένα πλήθος ιδιοτήτων του ερωτηματολόγιου μελετώνται, κατευθύνοντας τους ερευνητές προς την πιο **αξιόπιστη** και **έγκυρη** μεταφορά του στο νέο πληθυσμό.

Μια κλίμακα μέτρησης του άγχους για τη Στατιστική

Να δηλώσετε το βαθμό συμφωνίας ή διαφωνίας σας με τις παρακάτω προτάσεις (1-5): 1. Διαφωνώ απόλυτα 2. Διαφωνώ 3. Ούτε συμφωνώ/Ούτε διαφωνώ 4. Συμφωνώ 5. Συμφωνώ απόλυτα

1. Η Στατιστική με κάνει να θέλω να βάλω τα κλάματα
2. Η έννοια της τυπικής απόκλισης με ενθουσιάζει
3. Δεν ήμουν ποτέ καλός στα Μαθηματικά
4. Έχω ονειρευτεί τον Pearson να μου επιτίθεται με τον δείκτη συσχέτισής του
5. Δεν καταλαβαίνω τι έχει να μου προσφέρει η Στατιστική στη ζωή μου
6. Κλαίω με μαύρο δάκρυ όταν ακούω για δείκτες κεντρικής τάσης
7. Έχω ξυπνήσει κάτω απ' την κουβέρτα μου με τη σκέψη ότι είμαι παγιδευμένος σε μια καμπύλη κανονικής κατανομής
8. Ζαλίζομαι όταν βλέπω εξισώσεις
9. Αν είμαι άριστος στη Στατιστική, οι φίλοι μου θα με θεωρήσουν προβληματικό
10. Έχει τύχει να μην κοιμηθώ γιατί σκεφτόμουν την έννοια του μέσου όρου

Ψυχομετρικές Κλίμακες Μέτρησης

Ποιο/α εργαλείο/α έρευνας θα χρησιμοποιηθεί/ουν;

(α) Επιλογή από τα υπάρχοντα ψυχομετρικά εργαλεία
(βιβλιογραφία) **ή**

(β) Κατασκευή νέου ψυχομετρικού εργαλείου από τους ίδιους
τους ερευνητές

**Σε κάθε περίπτωση πρέπει να γίνει έλεγχος αξιοπιστίας και
εγκυρότητας των μετρήσεων που θα ληφθούν.**

Αξιοπιστία της Μέτρησης

- ❑ Σε όλους τους χώρους της επιστημονικής έρευνας όπου λαμβάνεται μία μέτρηση, τίθεται το ζήτημα της αξιοπιστίας της μέτρησης αυτής. Η αξιοπιστία αφορά **το κατά πόσο μια δοκιμασία αναδεικνύει το πραγματικό μέγεθος του υπό μέτρηση χαρακτηριστικού.**
- ❑ Για παράδειγμα, κατά πόσο μια ζυγαριά μετρά το πραγματικό βάρος ενός αντικειμένου ή μια κλίμακα ευφυΐας το βαθμό ευφυίας.
- ❑ Η αξιοπιστία είναι συνώνυμη της *συνέπειας* (*consistency*), όπως αυτή ορίζεται μέσω της *επαναληπτικότητας* (*repeatability*) και της *αναπαραγωγισμότητας* (*reproducibility*) ενός αποτελέσματος.
- ❑ Αν τα παραπάνω ισχύουν τότε μπορούμε να ισχυρισθούμε ότι μετράμε κατά ένα μεγάλο ποσοστό την πραγματική τιμή του βάρους, ενώ το ποσοστό των σφαλμάτων μέτρησης είναι μικρό. Δηλαδή μπορούμε να ισχυριστούμε ότι η μέτρηση μας είναι αξιόπιστη.

Αξιοπιστία της Μέτρησης

- Η εμφάνιση της έννοιας της αξιοπιστίας αποδίδεται στον Charles Spearman (1863 – 1945), ο οποίος πρώτος επισήμανε την ύπαρξη σφαλμάτων σε μια μέτρηση, που δεν ταυτίζονται με το δειγματοληπτικό σφάλμα.
- Ακόμη και σε περιπτώσεις όπου δεν υπάρχει δειγματοληψία, υπάρχουν σφάλματα. Πρόκειται για τα σφάλματα μέτρησης, τα οποία ταξινομούνται σε **τυχαία και συστηματικά**.

Πώς μετράμε την αξιοπιστία; (1)

Κλασσική Ψυχομετρική Θεωρία: Κάθε μέτρηση X που λαμβάνεται, είναι το άθροισμα δύο συνιστωσών, της πραγματικής τιμής T (true score) και του τυχαίου σφάλματος της μέτρησης ϵ . Δηλαδή ισχύει:

$$X = T + \epsilon \text{ ή διαφορετικά}$$

Παρατηρούμενη τιμή = πραγματική τιμή + σφάλμα

Τι ποσοστό όμως της κάθε μέτρησης είναι η πραγματική τιμή;
Αυτό ακριβώς καλείται να απαντήσει η έννοια της αξιοπιστίας.
Η αξιοπιστία ουσιαστικά είναι ο λόγος της πραγματικής τιμής προς την παρατηρούμενη σε μία μέτρηση, δηλαδή το πηλίκο

$$\text{αξιοπιστία} = \frac{\text{πραγματική τιμή}}{\text{παρατηρούμενη τιμή}} = \frac{T}{X} = \frac{T}{T + \epsilon}$$

Πώς μετράμε την αξιοπιστία; (2)

- Τεχνικά, η αξιοπιστία μέτρησης αφορά την διακύμανση των τιμών σε μία ομάδα ανθρώπων (και όχι τη μέτρηση σε μεμονωμένα άτομα). Συνεπώς έχει νόημα να μιλάμε για τη διακύμανση των τιμών (μιας μεταβλητής) σε ένα δείγμα ατόμων, ή την κατανομή των τιμών της εντός της ομάδας.

Συνεπώς:

$$\text{αξιοπιστία} = \frac{\text{Διακύμανση πραγματικών τιμών}}{\text{Διακύμανση παρατηρούμενων τιμών}}$$

- Ο υπολογισμός του παρονομαστή γίνεται στην έρευνα. Ο υπολογισμός της τιμής του αριθμητή είναι στην ουσία ανέφικτος. Με τη βοήθεια της στατιστικής (reliability analysis, SPSS) μπορούμε να εκτιμήσουμε το ποσοστό επί της συνολικής διακύμανσης που οφείλεται σε τυχαία σφάλματα μέτρησης. Μπορούμε λοιπόν να έχουμε μία εκτίμηση του ποσοστού της παρατηρούμενης διακύμανσης των σκορ που οφείλεται σε τυχαία σφάλματα μέτρησης, άρα και μία εκτίμηση του τυχαίου σφάλματος.

Δείκτες Αξιοπιστίας

- Οι δείκτες αξιοπιστίας παρέχουν μία εκτίμηση του ποσοστού της κοινής διακύμανσης μεταξύ του παρατηρούμενου και του πραγματικού σκορ.
- Ο πιο δημοφιλής
 - δείκτης εσωτερικής συνέπειας Cronbach's α (alpha), τιμές μεταξύ 0 και 1

Ενδεικτικές τιμές δείκτη αξιοπιστίας α:

< 0,6 η κλίμακα είναι αναξιόπιστη

0,6 το ελάχιστο αποδεκτό όριο (μή αποδεκτό για κλίμακες με πολλά items-προτάσεις)

0,7 επαρκές, αλλά όχι καλό

0,8 καλύτερο

0,95 πολύ υψηλή αξιοπιστία (μάλλον σπάνιο)

Έννοιες της αξιοπιστίας (1)

Υπάρχουν τρεις διαφορετικές έννοιες της αξιοπιστίας: η σταθερότητα, η ισοδυναμία και η εσωτερική συνέπεια.

A. Σταθερότητα (των μετρήσεων στο χρόνο)

Συνέπεια των μετρήσεων των ίδιων μεταβλητών με το ίδιο εργαλείο μέτρησης, στο ίδιο ή παρόμοιο δείγμα στο χρόνο. Προκειμένου να βεβαιωθούμε ότι το ύψος του τυχαίου σφάλματος είναι χαμηλό κάνουμε διαδοχικές μετρήσεις με τα ίδια όργανα μέτρησης στο ίδιο δείγμα ή σε δείγμα με παρόμοια χαρακτηριστικά.

Το διάστημα μεταξύ των μετρήσεων θα πρέπει να είναι αρκετά μεγάλο, ώστε να μην θυμάται το άτομο τις απαντήσεις του, αλλά όχι υπερβολικά μεγάλο, γιατί υπάρχει ο κίνδυνος να αλλάξει τη τιμή της μεταβλητής που μας ενδιαφέρει να μετρήσουμε (πχ λόγω ωρίμανσης ή μάθησης).

Προβλήματα προσέγγισης: - Πρακτικές δυσκολίες στην επανάληψη μετρήσεων - Πρόβλημα διακύμανσης των τιμών των μεταβλητών στο χρόνο λόγω εξάσκησης, μνήμης

Έννοιες της αξιοπιστίας (2)

Β. Ισοδυναμία (μορφών, βαθμολογητών)

Η ισοδυναμία μπορεί να επιτευχθεί μέσω της χρήσης ισοδύναμων μορφών μιας δοκιμασίας ή ενός οργάνου συλλογής δεδομένων. Στην περίπτωση που η ισοδύναμη μορφή επιφέρει παρόμοια αποτελέσματα, τότε μπορούμε να ισχυριστούμε ότι το εργαλείο επιδεικνύει τη συγκεκριμένη μορφή αξιοπιστίας. Η αξιοπιστία ως ισοδυναμία μπορεί επίσης να επιτευχθεί μέσω της συμφωνίας μεταξύ των ερευνητών που συμμετέχουν στην έρευνα.

Γ. Εσωτερική Συνέπεια (εργαλείου μέτρησης)

Ενώ οι προηγούμενες μορφές απαιτούν τα ερευνητικά εργαλεία να εφαρμοστούν δύο ή περισσότερες φορές, για να έχουμε εσωτερική συνέπεια απαιτείται να εφαρμοστούν μόνο μία φορά.

Μια εκτίμηση της εσωτερικής συνέπειας του ερευνητικού εργαλείου δίνει ο δείκτης Cronbach's α.

Είδη αξιοπιστίας (1)

1. Αξιοπιστία ελέγχου-επανελέγχου (test-retest reliability)

Αναφέρεται στο βαθμό συνέπειας που παρουσιάζουν διαδοχικές μετρήσεις με το ίδιο εργαλείο μέτρησης στο ίδιο δείγμα και κάτω από τις ίδιες συνθήκες. Προϋποθέτει τη διαχρονική σταθερότητα των μετρήσεων, Πρακτικές δυσκολίες στη διεξαγωγή διαδοχικών μετρήσεων, προβλήματα εξάσκησης, μνήμης.

2. Ισοδύναμων ή Παράλληλων Μορφών (parallel forms)

Διαδοχικές μετρήσεις με ισοδύναμες μορφές του ίδιου οργάνου, και υπολογισμός ενός δείκτη συσχέτισης των μετρήσεων. Αντιμετώπιση των επιδράσεων εξάσκησης ή μνήμης. Προϋπόθεση ότι και οι δύο μορφές του εργαλείου μετρούν το ίδιο χαρακτηριστικό. Χρησιμοποιείται όταν έχουμε μετρήσεις «Πριν και Μετά (Pre/Post)»

3. Αξιοπιστία μεταξύ παρατηρητών/βαθμολογητών (inter-rater reliability)

Ο βαθμός συνέπειας των εκτιμήσεων μεταξύ δύο βαθμολογητών ή δύο παρατηρητών (όταν πρόκειται για παρατήρηση).

Είδη αξιοπιστίας (2)

4. Τεχνική των δύο τμημάτων (split-half technique)

Υπολογισμός του βαθμού συνοχής δύο τμημάτων του ίδιου εργαλείου (συσχέτιση βαθμολογιών). Το εργαλείο (τεστ) χορηγείται μία φορά, και ο διαχωρισμός των τμημάτων γίνεται εκ των υστέρων από τους ερευνητές με αυθαίρετο τρόπο. Μειονέκτημα το ότι η τιμή του δείκτης συσχέτισης μεταξύ των δύο μερών εξαρτάται από την επιλογή των ερωτήσεων για το κάθε τμήμα.

5. Δείκτης εσωτερικής συνέπειας του Cronbach (Cronbach's α)

Ο δείκτης α, είναι ένας δείκτης της εσωτερικής συνέπειας ενός τεστ (internal consistency), και υπολογίζεται από τη συσχέτιση της κάθε ερώτησης του τεστ με την συνολική κλίμακα. Είναι από τις περισσότερο δημοφιλείς τεχνικές υπολογισμού της αξιοπιστίας (Cronbach, 1951).

Παράγοντες που μειώνουν την αξιοπιστία

- α. Λάθη στη συμπλήρωση του εργαλείου μέτρησης
- β. Απαντήσεις «στη τύχη» (guessing)
- γ. Πολύ μικρός αριθμός ερωτήσεων (> 3 ερωτήσεις ανά διάσταση)
- δ. Ασαφώς διατυπωμένες ερωτήσεις
- ε. Οδηγίες συμπλήρωσης που δεν είναι ακριβείς και ξεκάθαρες
- στ. Σφάλμα που οφείλεται στους συμμετέχοντες (π.χ κοινωνικά αποδεκτή συμπεριφορά, κούραση, πλήξη κλπ)
- ζ. Μεταβολές στις συνθήκες διεξαγωγής της έρευνας

Ψυχομετρικά Χαρακτηριστικά Κλιμάκων Μέτρησης

Αξιοπιστία

Εγκυρότητα

«Μια κλίμακα μέτρησης μπορεί να είναι αξιόπιστη, αλλά όχι έγκυρη. Αντίθετα, αν μια κλίμακα είναι έγκυρη τότε είναι και αξιόπιστη»

Βασικά Είδη Εγκυρότητας (Validity)

- Εσωτερική Εγκυρότητα (Internal validity)
- Εξωτερική Εγκυρότητα (External validity)
- Εγκυρότητα Περιεχομένου (Content validity)
- Εγκυρότητα Κριτηρίου (Criterion validity)
- Εννοιολογική Εγκυρότητα ή Εγκυρότητα Εννοιολογικής Κατασκευής (Construct validity)

Εσωτερική Εγκυρότητα

Εσωτερική εγκυρότητα: τα αποτελέσματα μιας μελέτης υποδεικνύουν μία αιτιώδη σχέση μεταξύ των μεταβλητών. Η ύπαρξη, όμως, συσχετίσεων δεν συνεπάγεται την ύπαρξη αιτιωδών σχέσεων. Μπορεί να μεσολαβούν λανθάνοντες παράγοντες που προκαλούν εμμέσως το αποτέλεσμα.

Επιδράσεις λανθανόντων παραγόντων:

- 1.Χρήση μόνο μιας ομάδας παρέμβασης
- 2.προβλήματα από τη χρήση πολλαπλών ομάδων...
- 3.αδυναμία ελέγχου εξωτερικών επιδράσεων στη συμπεριφορά και τις στάσεις των υποκειμένων (κοινωνικές επιδράσεις).

Για να αντιμετωπιστούν τα δύο πρώτα:

- α) χρήση ομάδας παρέμβασης / ομάδας ελέγχου,
- β) κατάλληλη επιλογή υποκειμένων για τις ομάδες (τυχαιοποίηση)
- γ) χειρισμοί πανομοιότυποι προς όλες τις ομάδες (με εξαίρεση τον χειρισμό της ανεξάρτητης μεταβλητής).

Εξωτερική Εγκυρότητα

Εξωτερική εγκυρότητα: τα αποτελέσματα να μπορούν αν γενικευθούν σε διαφορετικούς πληθυσμούς (διαφορετικά τμήματα του υπό μελέτη πληθυσμού) και καταστάσεις (διαφορετικές συνθήκες από αυτές που περιλαμβάνονται στην μελέτη).

Η δυνατότητα γενίκευσης είναι συνάρτηση του βαθμού ομοιότητας μεταξύ του δείγματος και των τμημάτων του πληθυσμού και μεταξύ των συνθηκών στις οποίες θέλουμε να γενικεύσουμε τα αποτελέσματα.

Εξαρτάται από: α) παράγοντες που σχετίζονται με τη δυνατότητα γενίκευσης σε άλλα άτομα ή ομάδες με διαφορετικά δημογραφικά χαρακτηριστικά, β) δυνατότητα γενίκευσης σε διαφορετικό κοινωνικό περίγυρο, γεωγραφική τοποθεσία ή κατάσταση, γ) παράγοντες που σχετίζονται με τη δυνατότητα γενίκευσης σε διαφορετικά χρονικά διαστήματα.

Εγκυρότητα Περιεχομένου

- ◆ Ο βαθμός στον οποίο τα μέρη του τεστ (item) εκπροσωπούν επάξια όλες τις πιθανές διαστάσεις του φαινομένου που αξιολογείται. Απαντά στο ερώτημα, *πόσο πλήρες είναι το εργαλείο μέτρησης;*
- ◆ Για παράδειγμα, μια κλίμακα κατάθλιψης δεν έχει εγκυρότητα περιεχομένου, όταν περιλαμβάνει μόνο τη συναισθηματική διάσταση και όχι τη διάσταση της συμπεριφοράς.

Διαδικασία εκτίμησης της εγκυρότητας περιεχομένου

- Εξ' ολοκλήρου υποκειμενική αξιολόγηση.
- Η αντιπροσωπευτικότητα των στοιχείων του εργαλείου μέτρησης αναφορικά με το αντικείμενο του ενδιαφέροντος γίνεται από τον ερευνητή, ενδεχομένως σε συνεργασία με συνεργάτες-κριτές.

1. Καταγραφή του συνόλου των στοιχείων που μπορεί να σχετίζονται με την έννοια που μας ενδιαφέρει. Η επιλογή γίνεται με βάση:

- α) τη θεματική ή εννοιολογική σχέση μεταξύ παρατηρήσιμων συμπεριφορών και του αντικειμένου μέτρησης
- β) εμπειρικά δεδομένα από προηγούμενες έρευνες που τεκμηριώνουν τη σχέση αυτή.

Διαδικασία εκτίμησης της εγκυρότητας περιεχομένου

2. από το σύνολο των στοιχείων αυτών επιλέγονται εκείνα που θεωρούνται ότι έχουν μεγαλύτερη συνάφεια με την υπό μελέτη ιδιότητα. Πιλοτική έρευνα σε δείγμα παρόμοιο με τον πληθυσμό του ενδιαφέροντος μπορεί να βοηθήσει να επιλεχθούν τα θέματα/προτάσεις που λειτουργούν με τον αναμενόμενο τρόπο και να απορριφθούν εκείνα που δεν είναι κατάλληλα.

Συμπερασματικά:

- η ακρίβεια του λειτουργικού προσδιορισμού θα είναι καταλυτική για τη αξιολόγηση της εγκυρότητας του περιεχομένου ενός εργαλείου/μιας δοκιμασίας μέτρησης.
- Η εγκυρότητα περιεχομένου καθορίζεται από την αρχή, από την επιλογή των θεμάτων/ερωτημάτων μέτρησης
- Η επιλογή των στοιχείων/θεμάτων είναι εξ ολοκλήρου υποκειμενική και βασίζεται στην υποκειμενική κρίση του ερευνητή/ερευνητών.

Εγκυρότητα Κριτηρίου

Ο βαθμός στον οποίο οι τιμές ενός τεστ σχετίζονται με τις τιμές κάποιου άλλου τεστ το οποίο θεωρείται κοινά αποδεκτό κριτήριο ή αντικειμενικό αποτέλεσμα.

Ένας τρόπος για να ελεγχθεί η εγκυρότητα βάσει κριτηρίου είναι να συσχετιστούν τα αποτελέσματα ενός τεστ με αυτά ενός μελλοντικού τεστ που μετρά την ίδια ή παρόμοια επίδοση ή συμπεριφορά. Αν το αρχικό τεστ μπορεί να προβλέψει μια μελλοντική επίδοση ή συμπεριφορά τότε υπάρχει προβλεπτική εγκυρότητα (predictive validity).

Ένας άλλος τρόπος είναι να συσχετιστούν οι μετρήσεις του τεστ με ένα άλλο που μετρά το ίδιο ή ένα συναφές χαρακτηριστικό. Αυτή είναι η συντρέχουσα εγκυρότητα (concurrent validity).

Διαδικασία προσδιορισμού της εγκυρότητας κριτηρίου

- επιλογή κατάλληλων συμπεριφορών που θα έχουν το ρόλο του κριτηρίου, και κατάλληλων εργαλείων για την μέτρησή τους
- επιλογή αντιπροσωπευτικού δείγματος συλλογή δεδομένων με το υπό μελέτη εργαλείο
- συλλογή δεδομένων με το εργαλείο-κριτήριο στην κατάλληλη χρονική στιγμή
- προσδιορισμός της σχέσης μεταξύ βαθμολογιών στο υπό μελέτη και το εργαλείο-κριτήριο

Διαδικασία προσδιορισμού της εγκυρότητας κριτηρίου

Προβλήματα:

- Η χρήση κατάλληλου εργαλείου-κριτηρίου (δύσκολος ο εντοπισμός κριτηρίων για προσδιορισμό εγκυρότητας, ιδιαίτερα για τις περισσότερο αφηρημένες μεταβλητές)
- Η χρησιμοποίηση μη επαρκούς δείγματος
- Μικρός βαθμός αξιοπιστίας των εργαλείων μέτρησης

Εννοιολογική Εγκυρότητα

- Η επιστημονική αξία των αποτελεσμάτων κάθε μελέτης που ασχολείται με μια εννοιολογική κατασκευή πχ. κατάθλιψη εξαρτάται με το αν το εργαλείο μέτρησης της κατάθλιψης μετρά αυτό και όχι κάτι άλλο (πχ. ψύχωση).
- Η εγκυρότητα εννοιολογικής κατασκευής **ενσωματώνει διαφορετικές μορφές εγκυρότητας** και απαντά στο ερώτημα, α) κατά πόσο το εργαλείο μέτρησης μετρά την έννοια του ενδιαφέροντος μας και β) κατά πόσο το εργαλείο μας είναι συναφές με παρόμοια εργαλεία ή παραλλαγές του ίδιου εργαλείου.
- Έχει να κάνει με τη διατύπωση υποθέσεων (σχετικά με το τι είναι η έννοια που μας ενδιαφέρει), που στηρίζονται σε μια θεωρία συμπεριφοράς ή διαμορφώνονται από την παρατήρηση μιας συγκεκριμένης συμπεριφοράς. Οι υποθέσεις αυτές μπορούν να επαληθευτούν ή να διαψευσθούν κατά την αξιολόγηση της εννοιολογικής εγκυρότητας. → απαραίτητη

Μέθοδοι εκτίμησης της εννοιολογικής εγκυρότητας

1. Ανάλυση παραγόντων (factor analysis)

- Μέθοδος της πολυμεταβλητής ανάλυσης με την οποία μπορούμε να ομαδοποιήσουμε ερωτήσεις οι οποίες μετρούν την ίδια εννοιολογική κατασκευή, έτσι ώστε να προσδιορίσουμε τελικά τις επίμερους διαστάσεις της. Έχει δύο μορφές, την επιβεβαιωτική (confirmatory) και τη διερευνητική (exploratory).

Π.χ. Έστω ότι ένας ερευνητής έχει χορηγήσει μια κλίμακα μέτρησης του σχολικού εκφοβισμού η οποία, σύμφωνα με τους κατασκευαστές της, περιέχει ερωτήσεις που αναφέρονται σε τρεις διαφορετικές διαστάσεις, το σωματικό, το λεκτικό και τον κοινωνικό εκφοβισμό. Αν η κλίμακα του σχ. εκφοβισμού μετρά όντως αυτές τις τρεις διαστάσεις (αν δηλαδή οι μετρήσεις έχουν εννοιολογική εγκυρότητα), τότε αναμένεται ότι οι απαντήσεις των υποκειμένων στις ερωτήσεις που ανήκουν στην ίδια διάσταση, θα συσχετίζονται περισσότερο μεταξύ τους απ' ότι με αυτές που ανήκουν σε διαφορετική διάσταση. Ο ερευνητής μπορεί να το ελέγξει αυτό με την επιβεβαιωτική ανάλυση παραγόντων.

Μέθοδοι εκτίμησης της εννοιολογικής εγκυρότητας

2. Πίνακας πολλαπλών ψυχομετρικών χαρακτηριστικών - πολλαπλών μεθόδων (multitrait-multimethod matrix)

- Ο ερευνητής μετρά περισσότερα του ενός ψυχομετρικά χαρακτηριστικά με περισσότερες της μιας μεθόδους έρευνας.
- Παράδειγμα. Έστω ότι μια ερευνητική ομάδα έχει κατασκευάσει ένα νέο εργαλείο για τη μέτρηση της κατάθλιψης σε εφήβους. Η ομάδα χορηγεί το εργαλείο σε ένα δείγμα εφήβων, μαζί με άλλα δύο διαδεδομένα εργαλεία μέτρησης της κατάθλιψης, και ένα εργαλείο μέτρησης του άγχους.
- Κάθε ένα από τα παραπάνω εργαλεία χορηγείται στο δείγμα με δύο διαφορετικές μεθόδους (π.χ. ημιδομημένη συνέντευξη και ερωτήσεις με κλίμακα τύπου Likert).

Μέθοδοι εκτίμησης της εννοιολογικής εγκυρότητας

Στη συνέχεια ο ερευνητής υπολογίζει δύο ομάδες δεικτών:

α) **δείκτες συγκλίνουσας εγκυρότητας (convergent validity).**

Υπολογίζει τους συντελεστές συσχέτισης μεταξύ των μετρήσεων της ίδιας εννοιολογικής κατασκευής (της κατάθλιψης) που χορηγήθηκε με τις δύο διαφορετικές μεθόδους. Αν η συγκλίνουσα εγκυρότητα των μετρήσεων είναι υψηλή, οι συσχετίσεις θα πρέπει να είναι υψηλές.

β) **δείκτες αποκλίνουσας ή διακριτικής εγκυρότητας (divergent or discriminant validity coefficients).**

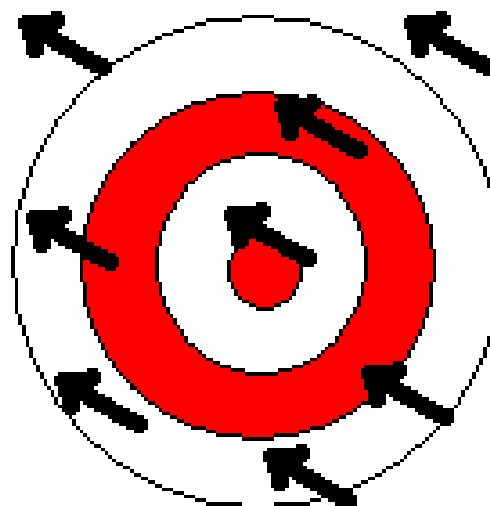
Υπολογίζει τους συντελεστές συσχέτισης μεταξύ των μετρήσεων των διαφορετικών εννοιολογικών κατασκευών που χορηγήθηκαν με την ίδια ή με διαφορετικές μεθόδους (π.χ. συσχέτιση αγχους και κατάθλιψης με συνέντευξη, άγχους με συνέντευξη και κατάθλιψης με ερωτήσεις τύπου Likert...) Αυτοί οι συντελεστές συσχέτισης θα πρέπει να είναι σχετικά χαμηλοί, αν η αποκλίνουσα εγκυρότητα των μετρήσεων είναι υψηλή.

Πιθανά Πρακτικά Προβλήματα κατά την Εκτίμηση της Αξιοπιστίας και της Εγκυρότητας

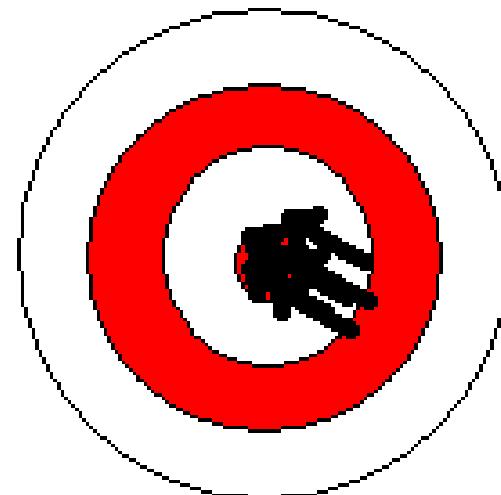
- 1. Μικρό μέγεθος δείγματος.** Για την εκτίμηση της αξιοπιστίας και της εγκυρότητας των μετρήσεων μιας κλίμακας απαιτούνται σχετικά μεγάλα δείγματα. Πρακτικός κανόνας: 5 με 10 υποκείμενα για κάθε ερώτηση μιας κλίμακας.
- 2. Χρήση μη αντιπροσωπευτικών και μη κατάλληλων δειγμάτων.** Π.χ. ένα ερωτηματολόγιο γνώσεων που κατασκευάστηκε για μαθητές γυμνασίου, μπορεί να μην είναι κατάλληλο για μαθητές δημοτικού.
- 3. Αν βρεθεί ότι υπάρχει εσωτερική συνέπεια, αυτό δε συνεπάγεται ότι υπάρχει και παραγοντική εγκυρότητα.** Κάποιες ερωτήσεις μπορεί να συσχετίζονται συμπτωματικά μεταξύ τους και συνεπώς να προκύπτει υψηλή τιμή του δείκτη α του Cronbach, αλλά αυτό δε σημαίνει ότι οι ερωτήσεις μετρούν μία μόνο διάσταση.
- 4. «Μαγείρεμα» των δεδομένων.** Προσπαθώντας να καταλήξουμε σε μία κλίμακα με υψηλή αξιοπιστία υπάρχει ο κίνδυνος να καταλήξουμε σε ένα ερωτηματολόγιο με πολύ λίγες ερωτήσεις. Όταν ενδιαφερόμαστε να κατασκευάσουμε μια κλίμακα είναι συνετό να έχουμε ένα θεωρητικό υπόβαθρο που να μας κατευθύνει, και να αποφεύγουμε το μαγείρευμα των δεδομένων.



Αξιόπιστο, Μη Έγκυρο



Μη Έγκυρο, Μη Αξιόπιστο



Έγκυρο, Αξιόπιστο