

Στοιχεία ψυχομετρίας

Δρ. Γεώργιος Δ. Γιαγλής (giaglis@hotmail.com)

Ορισμός

Ψυχομετρία (psychometrics) είναι η εφαρμοσμένη επιστήμη που μελετά τη μέτρηση ψυχολογικών χαρακτηριστικών (π.χ. γνώσεις, ικανότητες, στάσεις, προσωπικότητα, εκπαιδευτική επίδοση) με ερωτηματολόγια, δοκιμασίες (τεστ), κλίμακες κλπ.

Εναλλακτικά, μπορεί κανείς να πει ότι ψυχομετρία είναι

- ο κλάδος της ψυχολογίας που ασχολείται με το σχεδιασμό, χορήγηση και αξιολόγηση ποσοτικών δοκιμασιών που μετρούν ψυχολογικές μεταβλητές, όπως νοημοσύνη, ικανότητα και προσωπικότητα,
- η θεωρία και η τεχνική των νοητικών μετρήσεων (λεξικό Merriam-Webster),
- η ποσοτικοποίηση και ανάλυση των ατομικών διαφορών, που περιλαμβάνει την ανάπτυξη και την αξιολόγηση ψυχολογικών δοκιμασιών
- η συστηματική χρήση δοκιμασιών (τεστ) με σκοπό την ποσοτική μέτρηση των ψυχικών φαινομένων και της ανθρώπινης συμπεριφοράς (λεξικό Τριανταφυλλίδη)

Ιστορία

"Πατέρας" της ψυχομετρίας θεωρείται ο Francis Galton (1822-1911), ξάδερφος του Charles Darwin, που μεταξύ των ανθρωπομετρικών του μελετών συμπεριέλαβε και νοητικές δοκιμασίες. Σημαντική μορφή της ψυχομετρίας ήταν και ο James McKeen Cattell (1860-1944), ψυχολόγος, που προώθησε τη συστηματική χορήγηση νοητικών δοκιμασιών και ίδρυσε το πρώτο εργαστήριο ψυχομετρίας.

Η πρώτη εφαρμοσμένη ψυχομετρική δοκιμασία ήταν η κλίμακα νοημοσύνης των Binet και Simon (1905) που χρησιμοποιήθηκε σε μια προσπάθεια έγκαιρης αναγνώρισης των νοητικά υστερημένων παιδιών στο εκπαιδευτικό σύστημα της Γαλλίας. Αποτελείτο από 30 θέματα ποικίλης δυσκολίας, από το να αγγίξει κανείς τη μύτη ή το αυτί του μετά από αντίστοιχη εντολή μέχρι την ικανότητα να σχεδιάζει σκίτσα από μνήμης και να ορίζει αφηρημένες έννοιες. Με βάση την κλίμακα αυτή και σε συνάρτηση με τη χρονολογική ηλικία τους, τα παιδιά ταξινομούσαν σε κατηγορίες ανάλογα με το επίπεδο της νοημοσύνης τους.

Μέτρηση και ψυχομετρία

Όπως φαίνεται και από τον ορισμό που προαναφέρθηκε, η ψυχομετρία αναφέρεται στη μέτρηση ψυχολογικών, δηλαδή αφηρημένων, χαρακτηριστικών. Πώς όμως μπορεί κανείς να μετρήσει κάτι που είναι αφηρημένο, δηλαδή μη υλικό; Σύμφωνα με τον Stanley Smith Stevens (1946), μέτρηση είναι η απόδοση αριθμών σε αντικείμενα ή γεγονότα με βάση κάποιον κανόνα («*measurement, in the broader sense is defined as the assignment of numerals to objects or events according to rules*»). Ο καλύτερος κανόνας που

θα μπορούσε να βρεθεί σε αυτήν την περίπτωση είναι η σύγκριση με κάτι άλλο, παρόμοιο. Έτσι, με βάση τον Michell (1997) *μέτρηση είναι η εκτίμηση του μεγέθους μιας ποσότητας σε σύγκριση με κάποια άλλη.*

Αυτή η τακτική, δηλαδή ο ορισμός μιας ποσότητας συγκριτικά με κάποιαν άλλη, χρησιμοποιείται σε όλες τις επιστήμες. Για παράδειγμα, το βάρος ενός αντικειμένου μετράται συγκρίνοντάς το με το βάρος του Διεθνούς Πρότυπου Χιλιόγραμμου (International Prototype Kilogram), ενός κυλίνδρου από λευκόχρυσο και ιρίδιο, που διατηρείται σε αυστηρά σταθερές συνθήκες στο Διεθνές Γραφείο Βαρών και Μετρήσεων (International Bureau of Weights and Measurements) στη Γαλλία. Έτσι, εάν ένα αντικείμενο έχει διπλάσιο βάρος από το πρότυπο χιλιόγραμμο, τότε λέμε ότι έχει βάρος δύο κιλά. Εάν το βάρος του είναι τρεισήμισι φορές μεγαλύτερο, τότε θεωρούμε ότι ζυγίζει τρεισήμισι κιλά.

Μέτρηση ψυχολογικών εννοιών

Οι *ψυχολογικές έννοιες* (constructs) που μετρώνται από την ψυχομετρία αφορούν συνήθως κοινά χαρακτηριστικά των ατόμων (π.χ. γνώσεις, στάσεις, προσωπικότητα) που η εμπειρία μάς έχει δείξει ότι έχουν πρακτική σημασία στην κατανόηση και στην πρόβλεψη της συμπεριφοράς των ατόμων. Για παράδειγμα, μπορεί κανείς να πει ότι ένα από τα στοιχεία που χαρακτηρίζουν ένα άτομο είναι η αισιοδοξία ή απαισιοδοξία του. Εάν γνωρίζουμε ότι ένα άτομο είναι πολύ αισιόδοξο, θα μπορούμε να προβλέψουμε, μέχρι ενός σημείου, την αντίδρασή του απέναντι π.χ. σε μια δυσκολία. Επομένως η αισιοδοξία είναι μια ψυχολογική έννοια που θα μας ενδιέφερε να μετρήσουμε.

Το πρόβλημα με τις ψυχολογικές έννοιες είναι ότι είναι αφηρημένες και επομένως ο κάθε ερευνητής μπορεί να χρησιμοποιεί το δικό του ορισμό στη μελέτη τους. Εάν όμως ο καθένας χρησιμοποιεί διαφορετικό ορισμό, τα συμπεράσματα των μελετών δεν μπορούν να συγκριθούν ούτε είναι δυνατόν να «χτίσει» ένας επόμενος ερευνητής πάνω στα αποτελέσματα του προηγούμενου. Για το λόγο αυτό, είναι σημαντικό να ορίζεται μια ψυχολογική έννοια με έναν σαφή τρόπο, τόσο συγκεκριμένο, ώστε οι υπόλοιποι ερευνητές να μπορούν όχι μόνο να καταλάβουν αλλά και να επαναλάβουν μια έρευνα με τον ίδιο ακριβώς τρόπο.

Λειτουργικός ορισμός

Μια χρήσιμη ιδέα είναι αυτή του *λειτουργικού ορισμού* (operational definition, Bridgman, 1959), όπου μια έννοια ορίζεται με βάση τις διαδικασίες (λειτουργίες) που θα δείξουν ξεκάθαρα τη διαφορά μεταξύ δύο ποσοτήτων (ή, στην προκειμένη περίπτωση, ατόμων). Στον ορισμό π.χ. της «αριθμητικής ικανότητας» ο ερευνητής θα πρέπει να αποφασίσει (ορίσει) σε τι διαφέρει ένα άτομο που έχει αριθμητική ικανότητα από κάποιο που δεν έχει ή έχει λιγότερη. Δηλαδή, τι μπορεί να κάνει ο αριθμητικά περισσότερο ικανός από τον αριθμητικά λιγότερο ικανό. Ένας ερευνητής μπορεί, για παράδειγμα, να ορίσει ότι ο «αριθμητικά ικανός» γνωρίζει πώς να πραγματοποιεί τις τέσσερις βασικές αριθμητικές πράξεις (πρόσθεση, αφαίρεση, πολλαπλασιασμό, διαίρεση) σε τριψήφιους αριθμούς. Αυτός είναι ένας λειτουργικός ορισμός της αριθμητικής ικανότητας, που μπορεί εύκολα να μετρηθεί π.χ. με την κατασκευή μιας δοκιμασίας που περιέχει αυτές τις αριθμητικές πράξεις. Όσο περισσότερες σωστές απαντήσεις δώσει ένας εξεταζόμενος σε αυτή τη δοκιμασία, τόσο περισσότερη «αριθμητική ικανότητα» θα θεωρείται ότι έχει, με βάση αυτόν τον

ορισμό. Συμπερασματικά, ο λειτουργικός ορισμός είναι μια προσπάθεια για μετατροπή μιας ασαφούς έννοιας σε μια μετρήσιμη.

Στάδια στην κατασκευή μιας ψυχολογικής δοκιμασίας

Το πρώτο βήμα στην κατασκευή μιας δοκιμασίας που μετρά μια ψυχολογική έννοια είναι η διερεύνηση και συλλογή ενδεικτικών λειτουργιών, δηλαδή διαδικασιών, που θα χρησιμεύσουν στη μέτρηση της ψυχολογικής έννοιας. Για παράδειγμα, στην προσπάθεια κατασκευής μιας κλίμακας μέτρησης του «καλού φοιτητή» θα πρέπει κανείς να διερωτηθεί τι ορίζει τον «καλό» φοιτητή, τι κάνει ο «καλός» φοιτητής που δεν κάνει ο μη «καλός».

Στη συνέχεια, ο ερευνητής θα διαλέξει ορισμένες από αυτές τις διαδικασίες και πιο συγκεκριμένα τις πιο χρήσιμες, τις πιο ενδεικτικές, τις πιο σαφείς και τις πιο εύκολα μετρήσιμες, ώστε να σχηματίσει έναν λειτουργικό ορισμό της έννοιας. Με βάση αυτόν, θα κατασκευαστεί ένα αδρό εργαλείο, που θα μετρά τα στοιχεία του ορισμού αυτού.

Στο παράδειγμά μας, ο «καλός φοιτητής» μπορεί να οριστεί λειτουργικά ως αυτός που παίρνει καλούς βαθμούς, συμμετέχει στο μάθημα, τηρεί τις προθεσμίες στις εργασίες του, έρχεται εγκαίρως στις παραδόσεις, συνεργάζεται επιτυχώς με τους συμμαθητές του και επιδεικνύει σεβασμό στους διδάσκοντες. Επομένως, το ψυχομετρικό εργαλείο θα μετρά με συγκεκριμένο τρόπο τα πιο πάνω έξι χαρακτηριστικά που επιλέχθηκαν από τον ερευνητή.

Το επόμενο, βασικότατο βήμα, στη διαδικασία κατασκευής ενός ψυχομετρικού εργαλείου είναι ο *πιλοτικός έλεγχός* του. Ο ερευνητής θα πρέπει να το χρησιμοποιήσει για την αξιολόγηση ενός επαρκούς αριθμού φοιτητών, ώστε να διαπιστώσει το πώς λειτουργεί στην πράξη. Είναι κατανοητό ως εργαλείο; Είναι ξεκάθαρο; Μήπως έχει αμφίσημες έννοιες; Μήπως ξεχάστηκαν κάποια σημαντικά στοιχεία; Τα εμπειρικά δεδομένα (μετρήσεις) που θα προκύψουν από την πιλοτική έρευνα, ο ερευνητής θα τα χρησιμοποιήσει για να ελέγξει την αξιοπιστία και την εγκυρότητα του εργαλείου. Εάν τα αποτελέσματα δεν είναι ικανοποιητικά, ο ερευνητής θα προβεί σε διόρθωση του εργαλείου και θα επαναλάβει τον πιλοτικό του έλεγχο, μέχρι να μείνει ικανοποιημένος.

Στην πράξη, ποτέ ένα εργαλείο δεν είναι τέλειο αλλά και με την πάροδο των ετών η κουλτούρα των ανθρώπων (γλώσσα, γνώσεις, στάσεις, συμπεριφορές) αλλάζει. Αυτό έχει ως αποτέλεσμα όλα τα ψυχομετρικά εργαλεία να βρίσκονται σε μια διαδικασία διαρκούς εξέλιξης και αναθεώρησης (έρευνα → ψυχομετρική αξιολόγηση → διόρθωση → έρευνα... και ούτω καθεξής). Ένα χαρακτηριστικό παράδειγμα είναι η πιο διαδεδομένη κλίμακα μέτρησης της νοημοσύνης, η κλίμακα WAIS (Wechsler Adult Intelligence Scale), που ξεκίνησε το 1939 (ως Wechsler-Bellevue), αναθεωρήθηκε σε WAIS (1955), WAIS-R (1981), WAIS-III (1997) και, προς το παρόν, WAIS-IV (2008). Αξίζει να σημειωθεί ότι και το αρχικό εργαλείο (Wechsler-Bellevue) αποτελείτο από υποδοκιμασίες που υπήρχαν σε προϋπάρχοντα τεστ, όπως το Binet-Simon ή τα Alpha και Beta Intelligence Tests του Yerkes. Είναι, λοιπόν, προφανής η αέναη προσπάθεια για βελτίωση και διαρκή επικαιροποίηση του συγκεκριμένου ψυχομετρικού εργαλείου.

Ακρίβεια της μέτρησης

Καμία μέτρηση δεν είναι απόλυτα ακριβής. Για παράδειγμα, όταν λέμε «το ύψος μου είναι 1 μέτρο και 74 εκατοστά» δεν είμαστε σίγουροι αν το ύψος είναι ακριβώς 174 εκατοστά ή 174,3 εκατοστά ή 173,9 εκατοστά ή ακόμα και 174,3925... εκατοστά. Η διαφορά της πραγματικής τιμής από τη μετρούμενη τιμή ονομάζεται *λάθος* ή *σφάλμα μέτρησης* (measurement error). Δηλαδή:

$$\text{Λάθος μέτρησης} = \text{Πραγματική τιμή} - \text{Μετρούμενη τιμή}$$

Με άλλα λόγια, λάθος μέτρησης είναι κάθε απόκλιση από την πραγματική τιμή που οφείλεται στη διαδικασία μέτρησης.

Εννοείται ότι επιθυμούμε οι μετρήσεις μας να είναι όσο το δυνατόν πιο ακριβείς. Όταν στις μετρήσεις παρεισφρήσουν σφάλματα, οι συγκρίσεις είναι δύσκολες ή αδύνατες. Για παράδειγμα εάν η πραγματική «αριθμητική ικανότητα» των αγοριών είναι κάπου ανάμεσα στο 2 και στο 10, ενώ των κοριτσιών είναι ανάμεσα στο 4 και στο 8, δεν μπορούμε να είμαστε σίγουροι ποιο φύλο έχει μεγαλύτερη αριθμητική ικανότητα στο δείγμα μας. Εάν οι μετρήσεις μας είναι πιο ακριβείς και καταλήξουμε στο συμπέρασμα ότι η πραγματική «αριθμητική ικανότητα» των αγοριών είναι κάπου ανάμεσα στο 8 και στο 9, ενώ των κοριτσιών είναι ανάμεσα στο 6 και στο 7, τότε μπορούμε να βγάλουμε πιο ασφαλή συμπεράσματα για την υπεροχή ενός από τα δύο φύλα (και πάλι για το δείγμα μας).

Σε γενικές γραμμές τα σφάλματα μέτρησης χωρίζονται σε δυο κατηγορίες:

1. *Τυχαία σφάλματα* (random errors), που οφείλονται σε άγνωστες, αστάθμητες αιτίες. Τις περισσότερες φορές κατανέμονται ομοιόμορφα, δηλαδή δεν επηρεάζουν τη μέτρηση προς μια ορισμένη κατεύθυνση. Παράδειγμα τυχαίου σφάλματος παρουσιάζει μια παλιά, αναλογική ζυγαριά που ο δείκτης της ταλαντεύεται προτού σταματήσει και μερικές φορές δείχνει λίγο περισσότερο, ενώ άλλοτε δείχνει λίγο λιγότερο από την πραγματική τιμή.
2. *Συστηματικά σφάλματα* (systematic errors), που οφείλονται σε μια συγκεκριμένη αιτία και επηρεάζουν με συστηματικό και επαναλαμβανόμενο τρόπο τη μέτρηση. Παράδειγμα συστηματικού σφάλματος παρουσιάζει μια ζυγαριά που μονίμως μας δείχνει βαρύτερους από ό,τι είμαστε ή ένας αυστηρός καθηγητής που, διορθώνοντας τα γραπτά, δίνει μικρότερες βαθμολογίες από ό,τι μας αξίζουν.

Πηγές σφαλμάτων μέτρησης

Μερικές από τις αιτίες που είναι δυνατόν μια έρευνα να οδηγήσει σε εσφαλμένες μετρήσεις μιας ψυχολογικής έννοιας είναι οι εξής:

- **Σφάλμα δειγματοληψίας:** Αντί να χρησιμοποιήσει ένα αντιπροσωπευτικό δείγμα, ο ερευνητής έκανε μετρήσεις σε συγγενείς, φίλους, γνωστούς και επέλεξε βολικά και όχι τυχαία υποκείμενα. Επομένως, οι εκτιμήσεις της ψυχολογικής μεταβλητής που μετρήσε δεν μπορούν να γενικευθούν στον πληθυσμό.
- **Ανακριβή εργαλεία:** Το ψυχομετρικό εργαλείο που χρησιμοποιήθηκε δεν έχει την επιθυμητή ακρίβεια είτε επειδή δεν κατασκευάστηκε σωστά είτε επειδή έγινε κακή προσαρμογή του είτε

επειδή χρησιμοποιήθηκε για να μετρήσει κάτι διαφορετικό από αυτό που κατασκευάστηκε να μετράει. Επομένως, οι μετρήσεις που προέκυψαν από αυτό είναι αναξιόπιστες και ανακριβείς.

- Μεροληψία των υποκειμένων: Τα υποκείμενα δεν δίνουν ειλικρινείς απαντήσεις είτε επειδή θέλουν να φανούν καλύτερα από ό,τι είναι στην πραγματικότητα, είτε επειδή ντρέπονται και θέλουν να αποκρύψουν ένα πρόβλημά τους, είτε επειδή τους συμφέρει να το υπερτονίσουν, είτε επειδή απλά θέλουν να ευχαριστήσουν τον ερευνητή και απαντούν αυτό που νομίζουν ότι αυτός επιθυμεί.
- Μεροληψία του ερευνητή: Ο ερευνητής έχει μια ξεκάθαρη ιδέα για το τι περιμένει ότι θα προκύψει από την έρευνά του και, πιθανότατα, επιθυμεί για προσωπικούς λόγους ένα συγκεκριμένο συμπέρασμα. Έτσι, επηρεάζει υποσυνείδητα (ή και συνειδητά!) τα υποκείμενα της έρευνας και τους ωθεί να απαντήσουν προς την κατεύθυνση που εκείνος προσδοκά ή επιθυμεί.
- Εσφαλμένο/διαφορετικό επιστημολογικό πλαίσιο: Ο ερευνητής μελετά ένα φαινόμενο ξεκινώντας από εσφαλμένες ή απλά διαφορετικές αρχές από αυτές που χρειάζεται για να κατανοήσει πλήρως το φαινόμενο αυτό. Για παράδειγμα, προσπαθεί να μελετήσει τη σχολική επίδοση σε ένα σχολείο μιας διαφορετικής χώρας θεωρώντας ότι ο καλύτερος μαθητής είναι αυτός που απαντά σωστά στις περισσότερες ερωτήσεις που θέτει ο δάσκαλος στην τάξη. Αγνοεί όμως ότι στη συγκεκριμένη χώρα θεωρείται κοινωνικά μη αποδεκτό ένας μαθητής να απαντά σωστά μίαν ερώτηση, μπροστά σε όλους τους συμμαθητές του που δε γνωρίζουν την απάντηση. Με άλλα λόγια, ξεκινά τη μελέτη ενός φαινομένου θεωρώντας ότι «η ατομική επίδοση (του μαθητή) είναι σημαντικότερη από την επίδοση της ομάδας (της τάξης)», ενώ στη συγκεκριμένη χώρα ισχύει το αντίθετο, δηλαδή η ομαδική επίδοση θεωρείται σημαντικότερη και προστατεύεται έναντι του ατόμου. Αυτό έχει τελικά ως αποτέλεσμα οι μετρήσεις του για τη σχολική επίδοση των μαθητών να είναι εντελώς άστοχες.

Από όλες τις παραπάνω πηγές σφαλμάτων μέτρησης, η ψυχομετρία ασχολείται κατ' εξοχήν με τους τρόπους με τους οποίους μπορεί να περιοριστεί το σφάλμα που οφείλεται στα ανακριβή εργαλεία.

Αξιοπιστία

Όπως αναφέρθηκε πιο πάνω, ένα «καλό» εργαλείο είναι ένα εργαλείο που έχει μικρό σφάλμα μέτρησης. *Αξιοπιστία* (reliability) είναι η ιδιότητα ενός εργαλείου μέτρησης να δίνει την ίδια τιμή, όταν μετράει το ίδιο αντικείμενο. Επομένως, το αξιόπιστο εργαλείο μετρά πάντοτε με τον ίδιο (σχεδόν) τρόπο. Για παράδειγμα, μια ζυγαριά είναι αξιόπιστη, εάν κάθε φορά που ανεβαίνουμε μας δείχνει την ίδια τιμή βάρους.

Στην ψυχομετρία υπάρχουν τέσσερις διαφορετικές μέθοδοι, για να μελετήσει κανείς την αξιοπιστία ενός εργαλείου:

1. *Αξιοπιστία επαναληπτικών μετρήσεων* (test retest reliability). Σε αυτή τη μέθοδο χορηγούμε την ίδια δοκιμασία δύο (ή περισσότερες) φορές στα ίδια άτομα και στη συνέχεια συγκρίνουμε τις τιμές που βρήκαμε. Αυτή η μορφή αξιοπιστίας ονομάζεται και *σταθερότητα* (stability). Εάν μια δοκιμασία έχει αυτή τη μορφή αξιοπιστίας, τότε οι διαφορετικές μετρήσεις στα ίδια άτομα θα διαφέρουν μόνον λίγο (ή και καθόλου!) Ο χρόνος που μεσολαβεί ανάμεσα στις διαδοχικές χορηγήσεις της δοκιμασίας αποφασίζεται από τον ερευνητή και έχει σχέση και με τη μετρούμενη έννοια.
 - Από τη μια μεριά, το χρονικό διάστημα θα πρέπει να είναι αρκετά μικρό, ώστε οι συμμετέχοντες να μην έχουν προλάβει να «αλλάξουν» πολύ. Για παράδειγμα, εάν ένα εργαλείο που μετρά τα επίπεδα της κατάθλιψης το χορηγήσουμε δυο φορές με χρονική

απόσταση δύο μηνών, είναι πιθανόν να δούμε διαφορετικές τιμές, όχι επειδή το εργαλείο δεν είναι αξιόπιστο αλλά επειδή μέσα σε αυτό το χρονικό διάστημα έχει μεταβληθεί η συναισθηματική κατάσταση των υποκειμένων.

- Από την άλλη μεριά, το χρονικό διάστημα θα πρέπει να είναι αρκετά μεγάλο, ώστε οι συμμετέχοντες να μη θυμούνται τι απάντησαν την πρώτη φορά. Για παράδειγμα, εάν χορηγήσουμε μια δοκιμασία ξανά, πέντε λεπτά αργότερα από την πρώτη φορά, τότε είναι πιθανόν τα υποκείμενα να δώσουν απαντήσεις από μνήμης, παρά να ξανασκεφτούν τις απαντήσεις τους από την αρχή. Κι έτσι, θα καταλήξουμε να αξιολογήσουμε περισσότερο τη μνήμη των υποκειμένων, παρά την αξιοπιστία του εργαλείου μας.
2. *Αξιοπιστία παράλληλων μορφών* (alternate or parallel forms reliability). Σε αυτή τη μέθοδο κατασκευάζουμε δύο ισοδύναμες μορφές της ίδιας δοκιμασίας που έχουν παρόμοιο περιεχόμενο. Για παράδειγμα, στη δοκιμασία «αριθμητικής ικανότητας» χρησιμοποιούμε πολλαπλασιασμό τριψήφιων αριθμών αλλά χρησιμοποιούμε διαφορετικούς αριθμούς στην πρώτη (Α) και διαφορετικούς στη δεύτερη (Β) μορφή. Χορηγούμε και τις δυο μορφές στα ίδια άτομα και συγκρίνουμε τα αποτελέσματα. Θεωρητικά θα πρέπει οι δυο μορφές να δίνουν παρόμοια αποτελέσματα. Διαφορετικά θα πρέπει να καταλήξουμε στη χρήση πολύπλοκων στατιστικών μεθόδων που μπορούν να εξισώσουν τις βαθμολογίες των δυο δοκιμασιών.
- Σημαντικό επίσης είναι να χορηγηθούν αυτές οι δοκιμασίες εναλλάξ. Δηλαδή στα μισά υποκείμενα να χορηγηθεί πρώτα η μορφή Α και μετά η μορφή Β, ενώ στα υπόλοιπα μισά πρώτα η Β και μετά η Α. Με αυτόν τον τρόπο προσπαθούμε να εξισορροπήσουμε την επίδραση που έχει η χορήγηση της μιας μορφής στη χορήγηση της άλλης. Εάν π.χ. από λάθος όλα τα υποκείμενα συμπληρώσουν πρώτα τη μορφή Α και μετά τη Β και βρούμε υψηλότερες βαθμολογίες στη Β, αυτό μπορεί να σημαίνει ότι τα υποκείμενα «έμαθαν» από την Α, με αποτέλεσμα να έχουν υψηλότερη βαθμολογία στη Β, κι έτσι δε θα μπορούμε να τις συγκρίνουμε με δίκαιο τρόπο.
3. *Αξιοπιστία διχοτόμησης ή ημικλάστων* (split-half reliability). Αντί να κατασκευαστούν εναλλακτικές μορφές, είναι δυνατόν κάτω από κάποιες συνθήκες να χωρίσουμε τη δοκιμασία σε δυο ίσα τμήματα. Μπορούμε, δηλαδή, να θεωρήσουμε ως ένα τμήμα τα πρώτα μισά θέματα (ερωτήσεις) και ως δεύτερο τμήμα τα τελευταία μισά. Άλλος τρόπος είναι να θεωρήσουμε ως ένα τμήμα τα θέματα με περιττό (μονό) αριθμό και ως δεύτερο τμήμα τα θέματα με άρτιο (ζυγό) αριθμό. Σε αυτή τη μέθοδο χορηγούμε τη δοκιμασία στην ίδια ομάδα υποκειμένων και συγκρίνουμε τους βαθμούς τους στα δύο αυτά τμήματα.
- Μια βασική προϋπόθεση για την εφαρμογή αυτής της μεθόδου μελέτης της αξιοπιστίας είναι η δοκιμασία να είναι ομοιογενής, δηλαδή όλες οι ερωτήσεις να είναι παρόμοιες και ισοδύναμες ως προς τη σημασία τους. Διαφορετικά θα καταλήξει το ένα τμήμα να είναι διαφορετικό ή και σημαντικότερο από το άλλο.
 - Μια επέκταση αυτής της μορφής αξιοπιστίας είναι ο υπολογισμός του συντελεστή α του Cronbach. Ο συντελεστής αυτός αποτελεί ένα μέτρο της συσχέτισης όλων των θεμάτων μεταξύ τους και υψηλές τιμές του ερμηνεύονται συχνά ως ένδειξη ότι όλα τα θέματα μετρούν την ίδια ψυχολογική έννοια. Τότε η μορφή αυτή της αξιοπιστίας ονομάζεται και *εσωτερική συνέπεια* (internal consistency).
4. *Αξιοπιστία μεταξύ κριτών ή βαθμολογητών* (inter-rater reliability). Όλες σχεδόν οι ψυχομετρικές δοκιμασίες, εκτός ίσως από τις ερωτήσεις πολλαπλής επιλογής, απαιτούν την κρίση ενός εξεταστή για να βαθμολογηθούν. Για παράδειγμα, οι δοκιμασίες που απαιτούν προφορικές απαντήσεις, αντιγραφή ενός σχήματος, παράθεση ενός ορισμού κλπ απαιτούν από το βαθμολογητή να λάβει

κάποιες αποφάσεις. Αυτή είναι μια επιπρόσθετη πηγή λάθους μέτρησης, γιατί ένας κριτής μπορεί π.χ. να είναι επιεικής, ενώ κάποιος άλλος αυστηρός. Για να μελετήσουμε κατά πόσο μια δοκιμασία επηρεάζεται ή όχι από τέτοιους παράγοντες, τη χορηγούμε μία φορά και στη συνέχεια ζητούμε να τη βαθμολογήσουν δύο (ή περισσότεροι) κριτές. Εάν οι διάφοροι βαθμολογητές δίνουν την ίδια (σχεδόν) τιμή για το ίδιο υποκείμενο, τότε η δοκιμασία αυτή έχει υψηλή αξιοπιστία μεταξύ κριτών.

Παράγοντες που επηρεάζουν την αξιοπιστία ενός εργαλείου

1. Το μέγεθος μιας δοκιμασίας μπορεί να επηρεάσει αρνητικά την αξιοπιστία της, καθώς μακροσκελείς δοκιμασίες τείνουν να προκαλούν κόπωση και ανία στα υποκείμενα. Αυτό έχει ως αποτέλεσμα να απαντούν χωρίς προσοχή και άρα χωρίς ακρίβεια.
2. Η ομοιογένεια στη μορφή αλλά και στις ερωτήσεις είναι σημαντική. Για παράδειγμα, ένα εργαλείο που έχει ερωτήσεις Ναι/Όχι, ερωτήσεις με βαθμολόγηση σε 5-βαθμη και μετά σε 4-βαθμη κλίμακα Likert και μετά επιλογή της σημαντικότερης απάντησης από μια λίστα επιλογών είναι πιθανόν να δημιουργήσει κόπωση και σύγχυση στα υποκείμενα. Ένα ερωτηματολόγιο με ομοιόμορφες ερωτήσεις (π.χ. όλες Ναι/Όχι) είναι πιο εύκολο να απαντηθεί και άρα λιγότερο πιθανό να γίνουν λάθη ή απροσεξίες.
3. Συνεργασιμότητα των υποκειμένων. Είναι αυτονόητο ότι οι ερωτώμενοι θα πρέπει να έχουν ένα κίνητρο, για να συμμετέχουν στην αξιολόγηση με διάθεση και ειλικρίνεια αλλιώς θα δώσουν τυχαίες ή καθόλου απαντήσεις. Στα πιθανά κίνητρα συμπεριλαμβάνονται η αμοιβή (χρηματική ή συμβολική), η προσδοκία κέρδους (π.χ. λαχνός σε κλήρωση), το άμεσο θεραπευτικό όφελος (π.χ. δοκιμασίες που θα βοηθήσουν στη διάγνωση ή θα κατευθύνουν τη θεραπεία τους) ή το έμμεσο θεραπευτικό όφελος (π.χ. τυποποίηση δοκιμασιών που θα βοηθήσουν την επιστήμη και την υγεία) κλπ
4. Οι σταθερές ή παρόμοιες συνθήκες χορήγησης, π.χ. σε δωμάτιο ήσυχο, φωτεινό, όπου είναι παρόντες μόνο ο εξεταστής και ο εξεταζόμενος, θα οδηγήσουν σε αξιόπιστες βαθμολογίες. Αντίθετα, θορυβώδεις συνθήκες, διακοπή από εξωτερικούς παράγοντες ή χορήγηση π.χ. αργά το απόγευμα είναι δυνατόν να επηρεάσουν την επίδοση των συμμετεχόντων και να προκαλέσουν ανακρίβεια στις μετρήσεις.
5. Ο χορηγητής είναι ίσως η πιο σημαντική παράμετρος της χορήγησης. Οι οδηγίες χορήγησης μιας δοκιμασίας οφείλουν να είναι σαφείς και αυστηρές και να τηρούνται κατά γράμμα από το χορηγητή. Εάν π.χ. ένας χορηγητής είναι πέραν του δέοντος βοηθητικός ή, αντίθετα, δε δίνει τις απαραίτητες διευκρινήσεις, θα επηρεάσει τη βαθμολογία των υποκειμένων. Επειδή συχνά οι οδηγίες χορήγησης ενός ψυχομετρικού εργαλείου είναι πολύπλοκες και οι πιθανές καταστάσεις που μπορούν να προκύψουν σε μια χορήγηση είναι γενικά απρόβλεπτες, οι χορηγητές εκπαιδεύονται και πιστοποιούνται ως ικανοί για τη χορήγηση συγκεκριμένων ψυχομετρικών εργαλείων.
6. Αντίστοιχα ή και περισσότερο σημαντική είναι η διαδικασία της βαθμολόγησης. Το πρωτόκολλο (διαδικασία) αξιολόγησης μιας δοκιμασίας είναι ένας αυστηρά καθορισμένος τρόπος βαθμολόγησης που συχνά απαιτεί εκπαίδευση και εμπειρία.
7. Πέραν των ερωτήσεων πολλαπλής επιλογής, οι ψυχομετρικές δοκιμασίες έχουν ένα βαθμό αξιοπιστίας μεταξύ κριτών που σχεδόν ποτέ δεν είναι τέλειος. Για το λόγο αυτό, όσο λιγότεροι (και ίδιοι) είναι οι βαθμολογητές των δοκιμασιών, τόσο περισσότερο αξιόπιστα αναμένεται να είναι τα αποτελέσματα.

Εγκυρότητα

Εγκυρότητα (validity) είναι η ιδιότητα ενός εργαλείου να μετράει αυτό που σκοπεύει να μετρήσει. Για παράδειγμα, μια δοκιμασία μνήμης θα πρέπει να μετράει τη μνημονική ικανότητα και όχι π.χ. το εύρος των γενικών γνώσεων ενός υποκειμένου. Εννοείται ότι ένα εργαλείο με χαμηλή αξιοπιστία δεν μπορεί να έχει υψηλή εγκυρότητα, αφού λόγω της χαμηλής αξιοπιστίας του δεν μπορούμε να είμαστε σίγουροι για το τι μετράει ακριβώς! Οι κυριότερες μορφές εγκυρότητας ενός ψυχομετρικού εργαλείου είναι οι ακόλουθες:

1. *Εγκυρότητα περιεχομένου* (content validity). Αυτή η μορφή εγκυρότητας αφορά κυρίως τις δοκιμασίες εκπαιδευτικής αξιολόγησης και αναφέρεται στο κατά πόσο η δοκιμασία καλύπτει επαρκώς το εξεταζόμενο αντικείμενο. Για παράδειγμα, μια δοκιμασία «αριθμητικής ικανότητας» θα πρέπει να καλύπτει επαρκώς τις τέσσερις βασικές αριθμητικές πράξεις (πρόσθεση, αφαίρεση, πολλαπλασιασμό, διαίρεση). Επομένως μια δοκιμασία «αριθμητικής ικανότητας» που αποτελείται από 20 αριθμητικές πράξεις, 5 από κάθε είδος, μπορούμε να υποστηρίξουμε ότι έχει εγκυρότητα περιεχομένου.
 - ο Βασική προϋπόθεση για την αξιολόγηση της εγκυρότητας περιεχομένου μιας δοκιμασίας είναι ότι ένας ή περισσότεροι ειδικοί θα συμφωνήσουν ως προς το περιεχόμενο, τη δομή και τη σημασία των στοιχείων που συναποτελούν την έννοια. Στο παράδειγμά μας, θα πρέπει να συμφωνήσουν ότι η «αριθμητική ικανότητα» ορίζεται πλήρως από τις τέσσερις αριθμητικές πράξεις και ότι έχουν συγκριτικά την ίδια αξία.
2. *Φαινομενική εγκυρότητα* (face validity). Αυτή η μορφή εγκυρότητας απαντά στο ερώτημα κατά πόσο η δοκιμασία φαίνεται στον μη ειδικό ότι μετράει αυτό που πραγματικά μετράει. Παρότι θεωρητικά αυτή είναι η λιγότερο σημαντική μορφή εγκυρότητας, πρακτικά έχει σημασία στο βαθμό ότι βοηθάει τα υποκείμενα να δεχτούν καλύτερα την αξιολόγηση, να πειστούν ότι κάνουμε όντως αυτό που τους είπαμε ότι κάνουμε.
3. *Εγκυρότητα εξωτερικού κριτηρίου* (criterion validity). Αυτή η μέθοδος αξιολόγησης της εγκυρότητας στηρίζεται σε άλλα εργαλεία ή δοκιμασίες που συμφωνούμε ότι μετρούν την ίδια έννοια. Για παράδειγμα, ένα νεοσύστατο ερωτηματολόγιο κατάθλιψης δεχόμαστε ότι μετράει όντως την κατάθλιψη, εάν παρουσιάζει υψηλή συσχέτιση στις βαθμολογίες του με ένα άλλο ερωτηματολόγιο (ή γενικότερα μια άλλη αξιολόγηση) κατάθλιψης, που είναι γνωστό και αποδεδειγμένο ότι μετράει πραγματικά την κατάθλιψη. Αυτή η δεύτερη αξιολόγηση είναι το εξωτερικό κριτήριο και θεωρείται το βασικό πρότυπο (gold standard) μέτρησης της κατάθλιψης. Η εγκυρότητα εξωτερικού κριτηρίου διακρίνεται σε:
 - ο *Συγχρονική εγκυρότητα* (concurrent validity), όταν η συσχέτιση υπολογίζεται με άλλα εργαλεία που χορηγήθηκαν στους ίδιους συμμετέχοντες την ίδια περίπου χρονική περίοδο, και
 - ο *Προβλεπτική εγκυρότητα* (predictive validity), όταν το ψυχομετρικό εργαλείο συσχετίζεται και άρα προβλέπει τη βαθμολογία άλλων εργαλείων που χορηγούνται στους ίδιους συμμετέχοντες μετά από κάποιο ικανό χρονικό διάστημα.
4. *Εγκυρότητα εννοιολογικής κατασκευής* (construct validity). Αυτή η μέθοδος αξιολόγησης της εγκυρότητας εξετάζει κατά πόσο ο λειτουργικός ορισμός που χρησιμοποιείται από μια ψυχομετρική δοκιμασία για τη μέτρηση μιας ψυχολογικής έννοιας συμφωνεί με την τρέχουσα θεωρία για την έννοια αυτή. Από πρακτικής άποψης, η βαθμολογία στη δοκιμασία που μετράει την έννοια αυτή θα πρέπει:

- Να παρουσιάζει υψηλή συσχέτιση με δοκιμασίες που μετρούν παρόμοιες ή συγγενικές (με βάση τη θεωρία) έννοιες, δηλαδή να παρουσιάζει *συγκλίνουσα εγκυρότητα* (convergent validity), και
- Να παρουσιάζει χαμηλή ή καθόλου συσχέτιση με δοκιμασίες που μετρούν άσχετες (με βάση τη θεωρία) έννοιες, δηλαδή να παρουσιάζει *αποκλίνουσα εγκυρότητα* (divergent validity).

Είδη ψυχομετρικών εργαλείων και αξιολόγησης

Τα ψυχομετρικά εργαλεία μπορούν να κατηγοριοποιηθούν ανάλογα με το περιεχόμενο των ψυχολογικών εννοιών που μετρούν, ανάλογα με τον τρόπο χορήγησής τους και ανάλογα με τον τρόπο αξιολόγησης των αποτελεσμάτων.

Το περιεχόμενο των ψυχομετρικών εργαλείων μπορεί να αφορά:

- Την *προσωπικότητα*, όπως π.χ. το MMPI-II (Minnesota Multiphasic Personality Inventory), το TAT (Thematic Apperception Test), το NeoPI-R (NEO Personality Inventory - Revised), το EPQ (Eysenck Personality Questionnaire) κλπ
- Τη *νοημοσύνη*, όπως το WAIS (Wechsler Adult Intelligence Scale), το Raven Standard Progressive Matrices, η δοκιμασία Draw-A-Person των Goonenuough-Harris κλπ
- Τη *γενική ψυχοπαθολογία*, όπως το SCL-90R (Symptom Checklist 90 - Revised), το GHQ (General Health Questionnaire) κλπ
- Πιο ειδικά το *άγχος*, όπως το STAI (State-Trait Anxiety Inventory) και το BAI (Beck Anxiety Inventory) κλπ ή
- Την *κατάθλιψη*, όπως το BDI (Beck Depression Inventory), το HRSD (Hamilton Rating Scale for Depression), το ZSRDS (Zung Self-Rating Depression Scale), το CES-D (Center for Epidemiologic Studies Depression scale) κλπ
- *Νευροψυχολογικές* δοκιμασίες, όπως το MMSE (Mini-Mental State Examination), το TMT (Trail Making Test), το WCST (Wisconsin Card Sorting Test), το CVLT (California Verbal Learning Test), το RCFT (Rey Complex Figure Test) κλπ
- Τις *εκπαιδευτικές επιδόσεις*, όπως το SAT (Scholastic Aptitude Test), το GRE (Graduate Record Examinations), το TOEFL (Test Of English as a Foreign Language) κλπ
- Αλλά και μια πλειάδα από άλλες ψυχολογικές έννοιες, όπως ποιότητα ζωής, σεξουαλική λειτουργία, στυλ μάθησης, αυτο-αποτελεσματικότητα, προτίμηση χεριού και πάρα πολλές ακόμα.

Σε γενικές γραμμές ψυχομετρική ονομάζεται μια δοκιμασία όχι λόγω του περιεχομένου της (δηλαδή της ψυχολογικής έννοιας που μετρά) αλλά λόγω του τρόπου με τον οποίο έχει κατασκευαστεί ή/και βαθμολογείται.

Ανάλογα με το χρόνο που διατίθεται για την ολοκλήρωσή τους, οι ψυχομετρικές *δοκιμασίες ικανοτήτων* (ability or aptitude tests) διακρίνονται σε:

- *Δοκιμασίες ικανότητας* (power tests) στις οποίες τα θέματα είναι συνήθως πιο δύσκολα και πολύπλοκα αλλά υπάρχει επαρκής χρόνος για να ολοκληρωθεί η δοκιμασία στο σύνολό της, και

- *Δοκιμασίες ταχύτητας* (speed tests) στις οποίες τα θέματα ίσως να είναι πιο απλά αλλά υπάρχει περιορισμένος χρόνος για τη συμπλήρωση μεγάλου αριθμού θεμάτων.

Οι ψυχομετρικές ιδιότητες που έχει μια δοκιμασία μπορεί να διαφέρουν πολύ ανάλογα με το αν χορηγείται υπό πίεση χρόνου (ταχύτητα) ή με άπλετο χρόνο (ικανότητα). Επομένως ο τρόπος χορήγησης είναι πολύ σημαντικός. Εκτός από τις δοκιμασίες ικανοτήτων (π.χ. νοημοσύνης, νευροψυχολογικές, εκπαιδευτικής ή επαγγελματικής επάρκειας), η χορήγηση υπό πίεση χρόνου δε χρησιμοποιείται συνήθως σε άλλα είδη δοκιμασιών (π.χ. ψυχοπαθολογίας, προσωπικότητας κλπ).

Αξίζει να σημειωθεί ότι τις περισσότερες φορές κατά τη διάρκεια μιας δοκιμασίας παροτρύνουμε τον εξεταζόμενο και περιμένουμε από αυτόν να πετύχει την καλύτερη δυνατή επίδοση (*μέγιστη προσπάθεια*). Αυτός είναι και ο λόγος που είναι απαραίτητο, πριν από τέτοιου τύπου αξιολογήσεις να προηγείται ένα διάστημα εξοικείωσης του εξεταζόμενου με τον χώρο εξέτασης, καθώς και εγκαθίδρυσης μιας καλής σχέσης μεταξύ εξεταστή και εξεταζόμενου (rapport). Στις δοκιμασίες αυτές συνήθως αξιολογούμε το μέσο όρο των επιδόσεων του εξεταζόμενου (τα περισσότερα ψυχομετρικά εργαλεία), ενώ σπανίως μπορεί να λαμβάνεται υπόψη μόνον η καλύτερη επίδοση (π.χ. Longest Digit Span Forward, στη μνήμη αριθμών, ή στις αθλητικές δοκιμασίες) ή μόνον η χειρότερη επίδοση (π.χ. στις εξετάσεις οδήγησης). Σε αντιδιαστολή με τη μέγιστη προσπάθεια, η *συνήθης προσπάθεια* λαμβάνεται από πληροφορίες είτε αυτοαξιολόγησης (π.χ. ερωτηματολόγια) είτε από συγγενείς είτε από άμεση παρατήρηση του εξεταζόμενου στο περιβάλλον του.

Τέλος, ανάλογα με τον τρόπο αξιολόγησης της βαθμολογίας τα ψυχομετρικά εργαλεία μπορούν να διαχωριστούν (Glaser, 1963) σε αυτά που γίνονται:

- *Αξιολόγηση με κριτήριο* (criterion-referenced). Σε αυτόν τον τρόπο αξιολόγησης υπάρχει συνήθως ένα *όριο διαχωρισμού* (cut-off score). Τα άτομα που έχουν βαθμολογία πάνω από το διαχωριστικό όριο, θεωρείται ότι έχουν πετύχει το κριτήριο, δηλαδή ξέρουν όσα χρειάζεται να ξέρουν ή μπορούν να κάνουν όσα χρειάζεται να μπορούν να κάνουν. Συνήθως τα εργαλεία αυτά χρησιμοποιούνται στην εκπαιδευτική αξιολόγηση και χρησιμοποιούνται για να κρίνουν εάν π.χ. ένας μαθητής θα περάσει την τάξη, ένας φοιτητής θα πάρει πτυχίο ή ένας ιατρός θα κριθεί ικανός να ασκήσει την ιατρική (mastery test). Το διαχωριστικό όριο μπορεί να είναι αυθαίρετο ή να έχει επιλεγεί με κάποιον ειδικό τρόπο. Σημειωτέον ότι εδώ ο όρος *κριτήριο* δεν αναφέρεται στο όριο διαχωρισμού της βαθμολογίας αλλά στο αντικείμενο της κρίσης. Για παράδειγμα το κριτήριο μπορεί να είναι «Ο φοιτητής ξέρει επαρκώς τη θεωρία της ψυχομετρίας» και το διαχωριστικό όριο να είναι π.χ. να απαντήσει σωστά σε τουλάχιστον 60% των ερωτήσεων.
- *Αξιολόγηση με νόρμες* (norm-referenced). Σε αυτόν τον τρόπο αξιολόγησης η βαθμολογία ενός υποκειμένου δεν έχει σημασία αυτή καθαυτή αλλά μόνο συγκριτικά με τον πληθυσμό. Για παράδειγμα μια βαθμολογία σε δοκιμασία νοημοσύνης που αντιστοιχεί σε IQ 100 δε μας λέει σε πόσες ερωτήσεις ή δοκιμασίες απάντησε επιτυχώς ο εξεταζόμενος αλλά ότι πέτυχε βαθμολογία καλύτερη από το 50% του πληθυσμού (ατόμων ίδιου φύλου και ηλικίας). Με τον τρόπο αυτό μπορεί κανείς να αξιολογήσει τη σχετική σειρά κατάταξης ενός ατόμου στον πληθυσμό και να αξιολογήσει πόσο «σπάνια» είναι μια βαθμολογία. Ένα χαρακτηριστικό παράδειγμα είναι οι εξετάσεις εισαγωγής στα ανώτερα και ανώτατα εκπαιδευτικά ιδρύματα, όπου δεν έχει σημασία ο συνολικός βαθμός ενός υποψηφίου αλλά η επίδοσή του συγκριτικά με τους υπόλοιπους υποψηφίους. Ο όρος *νόρμες* αναφέρεται στο πόσο καλά τα πήγε ο πληθυσμός σε μια δοκιμασία και συνήθως απεικονίζονται σε ειδικούς πίνακες.

- *Ιδιοθελική αξιολόγηση (ipsative)*. Αυτός ο τρόπος αξιολόγησης συγκρίνει τη βαθμολογία ενός ατόμου σε σχέση με τον εαυτό του και όχι με άλλα άτομα. Αυτό μπορεί να πάρει τις εξής δύο μορφές:
 - Σύγκριση με παλαιότερες βαθμολογίες του ίδιου ατόμου στην ίδια δοκιμασία. Εφαρμόζεται κυρίως σε εκπαιδευτικό ή σε κλινικό περιβάλλον και αξιολογεί π.χ. τη βελτίωση του μαθητή μετά από μια συγκεκριμένη μέθοδο εκπαίδευσης ή τη βελτίωση ενός ασθενή μετά τη χορήγηση ενός συγκεκριμένου φαρμάκου.
 - Σύγκριση των απαντήσεων του ατόμου μεταξύ τους ή σύγκριση της βαθμολογίας του σε μια δοκιμασία με τις βαθμολογίες του σε άλλες δοκιμασίες. Με τον τρόπο αυτό μπορούν να αναγνωριστούν τα δυνατά ή αδύναμα σημεία ενός ατόμου ή να ανιχνευθούν οι προτιμήσεις του. Ως παραδείγματα αυτής της μεθόδου αναφέρονται α) η αξιολόγηση των δοκιμασιών επαγγελματικού προσανατολισμού, που κρίνουν εάν π.χ. ένας μαθητής προτιμά τα χειρωνακτικά από τα πνευματικά επαγγέλματα, και β) η αξιολόγηση εάν ένα παιδί έχει π.χ. μεγαλύτερη λεκτική (verbal) ή μη-λεκτική, πρακτική (performance) νοημοσύνη.

Δεκάλογος της καλής έρευνας (με ψυχομετρικά εργαλεία)

1. Σχεδιάστε σωστά το ερευνητικό εργαλείο, διατυπώστε ένα σαφή λειτουργικό ορισμό της ψυχολογικής έννοιας που θέλετε να μετρήσετε.
2. Χρησιμοποιήστε πολλαπλές ενδεικτικές ερωτήσεις (ή εργαλεία) για τις ασαφείς έννοιες. Έτσι θα είστε πιο σίγουροι ότι καλύπτετε όλες τις πλευρές του προβλήματος.
3. Διατυπώστε καθαρά τις ερωτήσεις σας. Σιγουρευτείτε ότι είναι κατανοητές από τα άτομα στα οποία απευθύνεται το ερωτηματολόγιο ή η δοκιμασία.
4. Κάντε μία τουλάχιστον πιλοτική έρευνα και με βάση τα αποτελέσματά της διορθώστε το ερευνητικό εργαλείο.

Σε κάθε περίπτωση, προτιμήστε ένα εργαλείο που έχει σταθμιστεί για τον ελληνικό πληθυσμό ή, κατ'ελάχιστον, έχει προσαρμοστεί για την ελληνική κουλτούρα. Η προσαρμογή θα περιλαμβάνει όχι μόνο γλωσσικές διαφορές αλλά και διαφορές στην εξοικείωση με έννοιες ή εικόνες, διαφορές στο νόημα αμφίσημων εκφράσεων ή εννοιών, διαφορές στο εκπαιδευτικό σύστημα μεταξύ χωρών κ.ο.κ.

5. Κατασκευάστε ένα ξεκάθαρο πρωτόκολλο έρευνας, που θα ορίζει με σαφήνεια τη μέθοδο ανεύρεσης του δείγματος, τη διαδικασία χορήγησης των δοκιμασιών (χρόνος, χώρος και τρόπος) αλλά και τη διαδικασία βαθμολόγησής τους.
6. Αυξήστε την ποιότητα (αλλά και την ποσότητα!) των δεδομένων σας,
 - φροντίζοντας ώστε το ερευνητικό σας πρωτόκολλο να είναι όσο το δυνατόν πιο σύντομο (για να μην κουράζει τους συμμετέχοντες),
 - παρέχοντας κατάλληλη αμοιβή ή/και κίνητρο στα υποκείμενα που θα συμμετάσχουν στην έρευνά σας,
 - μορφοποιώντας τις δοκιμασίες, ώστε να είναι ευδιάκριτες ακόμα και σε ηλικιωμένα άτομα με προβλήματα όρασης και ελκυστικές στην εμφάνιση (γραμματοσειρά, χρώμα, κουκίδες, πλαίσια)

- ο παρέχοντας έγκαιρη υπενθύμιση στα άτομα που θα επαναξιολογηθούν μετά από κάποιο χρονικό διάστημα.
7. Εκπαιδεύστε τους χορηγητές σας, ώστε να χορηγούν όπως πρέπει τις δοκιμασίες και να αντιδρούν με έναν σταθερό και ομοιόμορφο τρόπο σε όλες τις περιστάσεις που θα αντιμετωπίσουν κατά τη χορήγηση.
 8. Εφαρμόστε τυχαία δειγματοληψία, όπου αυτό είναι εφικτό. Διαφορετικά, μειώστε τουλάχιστον τη μεροληψία, αποφεύγοντας συγγενείς και φίλους ως υποκείμενα και ανευρίσκοντας συμμετέχοντες χρησιμοποιώντας όσο το δυνατόν περισσότερες και διαφορετικές μεθόδους προσέγγισης (γνωστοί γνωστών, αγγελία σε εφημερίδα, επίσκεψη σε ΚΑΠΗ, κλπ και κυρίως συνδυασμό πολλών από αυτές τις μεθόδους).
 9. Διασφαλίστε την ακρίβεια των δεδομένων σας, χρησιμοποιώντας σαφείς οδηγίες κωδικοποίησης (για το πέρασμα των δεδομένων σε μια ηλεκτρονική βάση με στόχο τη στατιστική ανάλυση) και επανειλημμένο έλεγχο της ορθότητας και της αληθοφάνειας των δεδομένων. Για παράδειγμα, σε μια κωδικοποίηση 1=άνδρας, 2=γυναίκα, ανιχνεύστε και διορθώστε πιθανές εσφαλμένες τιμές π.χ. 0, 3, 11,12 κ.ο.κ.
 10. Τέλος, ζητήστε βοήθεια (όποτε τη χρειάζεστε) από άλλους συναδέλφους ή εκπαιδευτές, όταν αντιμετωπίζετε προκλήσεις και προβλήματα που δεν είστε σίγουροι για το πώς να τα λύσετε.

Σε κάθε περίπτωση, η συλλογή δεδομένων υψηλής ποιότητας από ένα αντιπροσωπευτικό δείγμα είναι το πιο σημαντικό εχέγγυο για μια καλή έρευνα. Οι σύγχρονες και πολύπλοκες στατιστικές αναλύσεις δεν μπορούν ποτέ να διορθώσουν μια άσχημα σχεδιασμένη έρευνα με κακής ποιότητας δεδομένα.

Βιβλιογραφικές παραπομπές

Bridgman, P. W. (1959). *The way things are*. Cambridge: Harvard University Press.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18, 510–522.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–383.

Stevens, S.S. (1946). On the theory of scales and measurement. *Science*, 103, 677-680.

Προτεινόμενη βιβλιογραφία

Bryman, A., & Cramer, D. (2005). *Quantitative data analysis: A guide for social scientists*. Philadelphia, PA: Routledge.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: Wiley.

Βιβλία ψυχομετρίας στα Ελληνικά

Αλεξόπουλος, Δ. (1998). *Ψυχομετρία : Σχεδιασμός τεστ και ανάλυση ερωτήσεων*. Αθήνα: Ελληνικά Γράμματα.

Αλεξόπουλος, Δ. (2004). *Ψυχομετρία : Ιστορία, θεωρίες και γενικές αρχές*. Αθήνα: Ελληνικά Γράμματα.

Αλεξόπουλος, Δ. (2011). *Ψυχομετρία*. Αθήνα: Πεδίο.

Κοντοπούλου, Ε. (2002). *Ψυχομετρία*. Αθήνα : Interbooks, 2002.

Κουλάκογλου, Κ. (2002). *Ψυχομετρία και ψυχολογική αξιολόγηση*. Αθήνα : Εκδόσεις Παπαζήση.

Mellon, R. (2010). *Κλινική ψυχομετρία*. Αθήνα: Πεδίο.

Επίσης, στο συλλογικό έργο *Τα ψυχομετρικά εργαλεία στην Ελλάδα* των Α.Σταλικά, Σ.Τριλιβα, Π.Ρούσση (επιμ.) εκδόσεις Πεδίο (2012) υπάρχει κατάλογος με εργαλεία που έχουν κατασκευαστεί ή προσαρμοστεί για τον ελληνικό πληθυσμό, μαζί με ψυχομετρικά χαρακτηριστικά τους, βιβλιογραφικές παραπομπές και στοιχεία επικοινωνίας.