

Προβλήματα εγκυρότητας των ψυχολογικών μετρήσεων

ΑΡΙΣΤΟΤΕΛΗΣ ΚΑΝΤΑΣ

Πανεπιστήμιο Πατρών

ΠΕΡΙΛΗΨΗ

Η εγκυρότητα των ψυχολογικών μετρήσεων καθορίζεται από τη μια μεριά από την αξιοπιστία του προγνωστικού οργάνου και από την άλλη από την καταλληλότητα του κριτηρίου. Η σχετικότητα, η αρτιότητα και η μόλυνση του κριτηρίου είναι σημεία που πρέπει να προσέχονται κατά την επιλογή του κριτηρίου. Παράγοντες όπως το δειγματοληπτικό σφάλμα και ο περιορισμός του εύρους των μετρήσεων μπορούν να εμποδίσουν τη διαπίστωση της πραγματικής εγκυρότητας του προγνωστικού οργάνου. Με τη μέθοδο της μετα-ανάλυσης προϋπαρχουσών ερευνών είναι δυνατό να διαπιστωθεί η πραγματική εγκυρότητα του προγνωστικού οργάνου, πέρα από τις εκάστοτε καταστασιακές ιδιαιτερότητες.

Λέξεις κλειδιά: Γενίκευση εγκυρότητας, εγκυρότητα, μετα-ανάλυση.

Προγνωστικό όργανο και κριτήριο

Στο παρόν άρθρο θα αναφερθούμε σε ορισμένα προβλήματα που σχετίζονται με την εγκυρότητα των ψυχολογικών μετρήσεων. Τα θέματα που θα θίξουμε αναπτύχθηκαν ως προβληματική κυρίως στο χώρο της επιλογής προσωπικού αλλά αφορούν όλους τους τομείς της ψυχολογίας όπου επιχειρείται κάποια μορφή πρόβλεψης με βάση κάποιο προγνωστικό όργανο (ερωτηματολόγιο, τεστ κ.λπ.).

Προγνωστικό όργανο είναι κάθε μέτρηση με την οποία προσπαθούμε να προβλέψουμε μια μελλοντική συμπεριφορά (Κάντας, 1993). Ένα τεστ νοημοσύνης, για παράδειγμα, αποτελεί προγνωστικό όργανο με την έννοια ότι με βάση αυτό κάνουμε κάποιες προβλέψεις για τη συμπεριφορά του ατόμου στο σχολείο, στην εργασία, στην κοινωνία.

Για να εξακριβώσουμε την προγνωστική ικανότητα του οργάνου που χρησιμοποιούμε πρέπει να το συσχετίσουμε με την πραγματική συμπεριφορά, την οποία υποτίθεται ότι προβλέπει. Η συνάφεια ανάμεσα στις μετρήσεις που γίνονται με το προγνωστικό όργανο και τις μετρήσεις της πραγματικής συμπεριφοράς μας δείχνει το βαθμό εγκυρότητας του προγνωστικού μας οργάνου. Για να γίνει όμως κάτι τέτοιο είναι απαραίτητο να υπάρχει κάποια αξιόπιστη και έγκυρη μέτρηση της πραγματικής συμπεριφοράς. Η μέτρηση αυτή στην ψυχομετρική θεωρία ονομάζεται κριτήριο. Επειδή όμως η πραγματική συμπεριφορά (π.χ., επίδοση στο σχολείο ή στην εργασία) έχει πολλές όψεις, το κριτήριο που υιοθετούμε (π.χ., αξιολογήσεις προϊσταμένων ή μέση βαθμολογία) είναι αναγκαστικά ελλιπές, αποτελεί φτωχό αντίκρουσμα της πραγματικής συμπεριφοράς. Κατά συνέπεια δεν μπορούμε να καθορίσουμε με

ακρίβεια την εγκυρότητα, ή την προγνωστική ικανότητα των μετρήσεων που κάνουμε με το προγνωστικό όργανο. Το ίδιο, βέβαια, συμβαίνει και στην περίπτωση όπου η αξιοπιστία του προγνωστικού οργάνου είναι χαμηλή. Ένα τεστ για το άγχος, για παράδειγμα, δεν αποτελεί από μόνο του έγκυρη μέτρηση του άγχους, αν δεν έχει προηγουμένως αποδειχτεί ότι έχει υψηλή συνάφεια με το πραγματικό εκδηλούμενο άγχος. Η δυσκολία στην περίπτωση αυτή είναι να καθορίσουμε με ποιους φυσιολογικούς δείκτες, ή με ποιες συμπεριφορικές εκδηλώσεις, μπορούμε να μετρήσουμε το πραγματικό άγχος, ούτως ώστε να συσχετίσουμε τις μετρήσεις αυτές με τη μέτρηση που μας δίνει το ερωτηματολόγιο. Είναι φανερό ότι στην περίπτωση αυτή αναφερόμαστε στην εγκυρότητα σε σχέση με το κριτήριο. Αν ο προβληματισμός μας αναφέρεται σε αυτή καθαυτή την έννοια του άγχους και πώς αυτό ορίζεται, τότε αναφερόμαστε στην εγκυρότητα εννοιολογικής κατασκευής.

Για να διαπιστώσουμε επομένως την εγκυρότητα ενός προγνωστικού οργάνου, απαραίτητο είναι προηγουμένως να ορίσουμε το κριτήριο με το οποίο στη συνέχεια θα το συσχετίσουμε. Κατά την επιλογή του κριτηρίου είναι απαραίτητο να προσέχουμε διάφορα σημεία. Η *σχετικότητα του κριτηρίου*, το πόσο σχετικό δηλαδή είναι το κριτήριο με την πραγματική συμπεριφορά, είναι ένα πρώτο σημείο που πρέπει να μας απασχολήσει κατά την επιλογή του κριτηρίου. Οι βαθμοί, για παράδειγμα, έχουν σαφώς σχέση με τη σχολική επίδοση, ισχύει όμως το ίδιο και για τη διαγωγή. Οι αξιολογήσεις των προϊσταμένων έχουν σαφώς σχέση με την επίδοση στην εργασία. Ισχύει όμως το ίδιο και για την εμφάνιση. Σε μερικές περιπτώσεις η εμφάνιση έχει σχέση με την εργασία σε άλλες όμως δεν έχει. Ήδη από παλιά είχε τεθεί το ερώτημα του κατά πόσο η μέτρηση της νοημοσύνης πρέπει να συσχετίζεται με τη σχολική επίδοση. Έτσι, για τον καθορισμό της εγκυρότητας ενός προγνωστικού οργάνου, απαραίτητο είναι να επιλέξουμε ένα ή περισσότερα σχετικά κριτήρια.

Η *αρτιότητα του κριτηρίου* είναι το δεύτερο σημείο που χρειάζεται προσοχή. Έχουν ληφθεί υπόψη κατά τον ορισμό ή την κατάρτιση του κρι-

τηρίου όλες οι σχετικές συμπεριφορές ή αυτό είναι ελλιπές; Η εμφάνιση του ατόμου, για παράδειγμα, και η εν γένει συμπεριφορά σε ορισμένες εργασίες έχουν σημαντικό ρόλο και, αν δε λαμβάνεται υπόψη, το κριτήριο που χρησιμοποιούμε σε μια παρόμοια περίπτωση, είναι πιθανόν ελλιπές. Για να δώσουμε ένα ακόμη παράδειγμα: αν χρησιμοποιούμε ως κριτήριο της επίδοσης των μεταπτυχιακών φοιτητών τον αριθμό ή την ποιότητα των γραπτών βιβλιογραφικών διερευνήσεων, τότε έχουμε σχετικό αλλά ίσως ανεπαρκές κριτήριο, γιατί δεν αξιολογούμε πιθανόν την ερευνητική τους ικανότητα.

Η *μόλυνση του κριτηρίου* είναι ο τρίτος παράγοντας που πρέπει να προσέχουμε κατά την επιλογή, κατάρτιση ή βαθμολόγηση του κριτηρίου. Στην περίπτωση αυτή αναφερόμαστε στις τυχόν διαστρεβλώσεις που υφίσταται το κριτήριο από εξωτερικούς παράγοντες. Αν αυτός που αξιολογεί (βάζει δηλαδή κάποια βαθμολογία που θα χρησιμοποιηθεί ως κριτήριο) επηρεάζεται στην κρίση του από παράγοντες όπως, π.χ., η γνώση της βαθμολογίας άλλων κριτών, ή αν η επίδοση των κρινομένων επηρεάζεται από άλλους παράγοντες (π.χ., κοινωνικο-οικονομικούς), είναι φανερό ότι το κριτήριο που θα χρησιμοποιήσουμε έχει μολυνθεί από εξωτερικούς παράγοντες.

Ένας τρόπος αποφυγής των προβλημάτων που δημιουργούνται από το κριτήριο είναι η επιλογή πολλαπλών ή σύνθετων κριτηρίων. Το ερώτημα που τίθεται σε παρόμοιες περιπτώσεις είναι ποια θα είναι η *συνεισφορά* ή η *βαρύτητα* του κάθε επιμέρους κριτηρίου στο τελικό κριτήριο. Ένας δείκτης κεντρικής τάσης (π.χ., ένας μέσος όρος βαθμολογίας διάφορων μαθημάτων) επικαλύπτει πολλά επιμέρους ποιοτικά και ποσοτικά προβλήματα. Έχουν όλα τα μαθήματα την ίδια βαρύτητα; Μήπως η υψηλή συσχέτιση μεταξύ μερικών εξ αυτών διαστρεβλώνει το μέσο όρο; Όπως φέρεται να είπε χαρακτηριστικά ο Cattell: Δέκα άντρες και δυο μπουκάλια μπύρας δεν μπορούν να προστεθούν και να δώσουν το ίδιο άθροισμα όπως δύο άντρες και δέκα μπουκάλια μπύρας!

Παράγοντες μείωσης της τιμής του δείκτη εγκυρότητας

Πέρα όμως από τη χαμηλή αξιοπιστία στο προγνωστικό όργανο ή στη μέτρηση του κριτηρίου, υπάρχουν ακόμα δύο παράγοντες που μπορούν να λειτουργήσουν αρνητικά κατά την προσπάθεια διαπίστωσης της εγκυρότητας: το δειγματοληπτικό σφάλμα και ο περιορισμός εύρους. Το δειγματοληπτικό σφάλμα είναι αναπόφευκτο τη στιγμή που εργαζόμαστε με μικρά δείγματα. Ο περιορισμός εύρους αναφέρεται στο γεγονός ότι συνήθως στα δείγματά μας δεν εμφανίζεται όλο το δυνατό εύρος τιμών (π.χ., όταν μετράμε τη νοημοσύνη σε μαθητές λυκείου ή φοιτητές είναι σχεδόν αδύνατο να βρούμε πολύ χαμηλές τιμές).

Όλα τα παραπάνω έχουν ως αποτέλεσμα οι δείκτες αξιοπιστίας που συνήθως βρίσκουμε να μην είναι συνήθως οι πραγματικοί και μάλιστα σε μερικές περιπτώσεις να πηγαίνουν προς την αντίθετη κατεύθυνση, δηλαδή να είναι αρνητικοί, ενώ αναμένονται να είναι θετικοί. Όπως θα αναπτυχθεί πιο κάτω, με τη μέθοδο της μετα-ανάλυσης σήμερα μπορούμε να βρούμε τους πραγματικούς δείκτες που ισχύουν για τον πληθυσμό, αλλά για να γίνει κατανοητό το πρόβλημα θα δώσουμε ένα παράδειγμα (Hunter & Schmidt, 1989). Η μέθοδος της μετα-ανάλυσης έχει δείξει ότι η πραγματική εγκυρότητα (σχέση) μεταξύ γενικής γνωστικής ικανότητας και επίδοσης στην εργασία είναι του επιπέδου 0.75, μια πραγματικά υψηλή σχέση. Όταν κάνουμε μια σχετική έρευνα με κάποιο δείγμα εργαζομένων, το πιθανότερο είναι να βρούμε μια σχέση του επιπέδου 0.20 ή χαμηλότερη.

Η διαφορά αυτή οφείλεται στους ακόλουθους λόγους. Το γεγονός ότι ως κριτήριο της εργασιακής επίδοσης είναι πολύ πιθανό να χρησιμοποιήσουμε τις αξιολογήσεις των προϊσταμένων, που υπόκεινται σε σφάλμα (ως υποκειμενικοί τρόποι αξιολόγησης), κατεβάζει την εγκυρότητα στο 0.47. Και αυτό εφόσον χρησιμοποιηθούν οι αξιολογήσεις δύο προϊσταμένων του κάθε εργαζομένου, γιατί αν χρησιμοποιηθεί η αξιολόγηση ενός μόνο προϊσταμένου ο δείκτης κατεβαίνει ακόμα περισσότερο. Αν πάλι η αξιολόγηση των προϊσταμένων έγινε με βάση μια συνολική

κλίμακα και όχι με επιμέρους κλίμακες, που πιάνουν πολλές όψεις της εργασιακής συμπεριφοράς του εργαζομένου, ο δείκτης κατεβαίνει στο 0.32. Η αξιοπιστία του προγνωστικού οργάνου (π.χ., τεστ γνωστικών ικανοτήτων) είναι συνήθως του επιπέδου 0.80. Αυτό κατεβάζει το δείκτη εγκυρότητας στο 0.29. Ο περιορισμός εύρους που έγκειται, όπως ήδη είπαμε, στο γεγονός ότι αυτοί που αξιολογήθηκαν με βάση το τεστ δεν αποτελούν δείγμα που μπορεί να δώσει όλες τις πιθανές τιμές στο τεστ, κατεβάζει την εγκυρότητα στο 0.20. Αν προσθέσουμε και το δειγματοληπτικό σφάλμα ο δείκτης εγκυρότητας κατεβαίνει ακόμα πιο κάτω. Ενώ δηλαδή η πραγματική σχέση ανάμεσα στο προγνωστικό μας όργανο (τεστ) και το κριτήριο (επίδοση στην εργασία) είναι του επιπέδου 0.70, εμείς είναι πολύ πιθανό να βρούμε πολύ χαμηλή σχέση, ή καθόλου σχέση. Αν κάναμε αυτή τη μελέτη που περιγράψαμε με διάφορα δείγματα του επιπέδου $N=68$, οι μισοί δείκτες εγκυρότητας θα κυμαίνονταν μεταξύ 0.11 και 0.29, το 5% των δεικτών θα έχει τιμές άνω του 0.40, ενώ ένα άλλο 5% των δεικτών θα ήταν αρνητικοί.

Σύμφωνα με τον Schmidt (1992) οι έρευνες που στηρίζονται στον έλεγχο της μηδενικής υπόθεσης (H_0) ελέγχουν το σφάλμα Τύπου I (αν φαίνεται να υπάρχει σχέση, ενώ δεν υπάρχει) αλλά δεν ελέγχουν επαρκώς το σφάλμα Τύπου II (αν φαίνεται να μην υπάρχει σχέση, ενώ υπάρχει). Αν υπάρχει μια αληθής σχέση στον πληθυσμό, δηλαδή η μηδενική υπόθεση είναι μη αληθής στην πραγματικότητα, δεν είναι δυνατό να υπάρξει σφάλμα Τύπου I (δηλαδή να συμπεράνουμε εσφαλμένα ότι υπάρχει σχέση). Το μόνο σφάλμα που μπορεί να γίνει είναι Τύπου II, δηλαδή να μη βρούμε την υπάρχουσα σχέση. Το σφάλμα αυτού του τύπου δεν ελέγχεται επαρκώς στις έρευνες, όπως αναφέρθηκε παραπάνω.

Μετα-ανάλυση και γενίκευση εγκυρότητας

Τα προβλήματα αυτά λύνονται με τη μέθοδο της μετα-ανάλυσης (Hunter & Hunter, 1984. Hunter & Schmidt, 1989. Hunter & Schmidt, 1990. Hunter, Schmidt, & Jackson, 1982), κατά την

οποία μέσα από πολλές ομοειδείς έρευνες βρίσκουμε την πραγματική σχέση στον πληθυσμό. Παλαιότερα, οι επισκοπήσεις των υπάρχουσών ερευνών στηρίζονταν στην ερμηνεία που έδινε αυτός που έκανε την επισκόπηση στα ευρήματα. Με βάση τα προηγούμενα ήταν αναπόφευκτο μερικές έρευνες να πηγαίνουν προς τη μια κατεύθυνση και, π.χ., να στηρίζουν την ύπαρξη μιας σχέσης, ενώ μερικές άλλες να πηγαίνουν προς την άλλη κατεύθυνση και να δείχνουν πως η σχέση αυτή δεν υπάρχει. Οι ερευνητές που έκαναν τις επισκοπήσεις έβγαζαν ως εκ τούτου συνήθως το συμπέρασμα ότι υπάρχει καταστασιακή ιδιαιτερότητα, ότι δηλαδή μια σχέση σε ορισμένες περιπτώσεις ισχύει, ενώ σε άλλες δεν ισχύει, λόγω ιδιοτυπιών στο δείγμα ή στη συγκεκριμένη περίπτωση (Murphy, 1994). Με τη μετα-ανάλυση, και τη συγκεκριμένη μορφή της που ονομάζεται γενίκευση εγκυρότητας, μπορούμε πλέον να βρούμε αν μια σχέση πραγματικά υπόκειται σε διακυμάνσεις λόγω καταστασιακών ιδιαιτεροτήτων ή αν οι εμφανιζόμενες διακυμάνσεις είναι αποτέλεσμα στατιστικών διαστρεβλώσεων λόγω ανεπαρκούς δειγματοληψίας ή λόγω ελλιπούς εύρους στη διακύμανση των τιμών. Αν βρεθεί να υπάρχει μια πραγματική σχέση, ανεξάρτητη από τις συγκεκριμένες καταστάσεις, τότε έχουμε γενίκευση της εγκυρότητας. Έχει βρεθεί, για παράδειγμα, ότι οι γνωστικές ικανότητες σχετίζονται με την επίδοση στην εργασία, άσχετα από το είδος της εργασίας.

Η μέθοδος της μετα-ανάλυσης είναι κάτι περισσότερο από στατιστική μέθοδος. Σύμφωνα με κάποιους εκ των θεμελιωτών της στο χώρο της ψυχολογίας, οι επιπτώσεις της θα είναι ευρύτερες (Schmidt, 1992). Η μετα-ανάλυση έδειξε πόσο λίγες είναι οι πληροφορίες που υπάρχουν σε μια και μοναδική μελέτη. Έχει δείξει ότι, αντίθετα με τη διαδομένη αντίληψη, καμία πρωτογενής μοναδική μελέτη δεν μπορεί να λύσει ένα πρόβλημα ή να δώσει απάντηση σε κάποιο ερώτημα. Αν λάβουμε υπόψη τις αρχές της μετα-ανάλυσης, αυτό που διαφαίνεται είναι ότι υπάρχει ένα ισχυρό δόγμα εμπειρισμού, με υπερβολική αυτοπεποίθηση, στις επιστήμες της συμπεριφοράς και τις κοινωνικές επιστήμες, μια υπερβολική πίστη, δηλαδή, στα δεδομένα ως άμεση πηγή της

επιστημονικής αλήθειας και παράλληλα ανεπαρκής συνειδητοποίηση του πόσο παραπλανητικά μπορούν να είναι τα περισσότερα δεδομένα των κοινωνικών επιστημών, αν γίνουν αποδεκτά με βάση τη φαινομενική τους αξία, ή αν ερμηνευθούν απλοϊκά. Η ευρέως διαδομένη πεποίθηση ότι η έρευνα θα προοδεύσει αν αφήσουμε τα δεδομένα να μιλήσουν, δυστυχώς είναι λανθασμένη... Κατά συνέπεια μια μεμονωμένη μελέτη πρέπει να θεωρείται ένα απλό δεδομένο που θα συνεισφέρει σε μια μελλοντική μετα-ανάλυση. Έτσι το επιστημονικό κύρος και η αξία μιας μοναδικής μελέτης είναι κατά συνέπεια μειωμένα (Schmidt, 1992, σ. 1178).

Ο ίδιος συγγραφέας θέτει μάλιστα και έναν ιδιότυπο προβληματισμό: μέχρι τώρα η επιστημονική καταξίωση ερχόταν με την πρωτογενή έρευνα, ενώ τώρα διαφαίνεται ότι οι μετα-ανάλυσεις προσφέρουν μερικές φορές περισσότερα στη θεωρία από ό,τι η πρωτογενής έρευνα, πράγμα που μπορεί να ωθήσει πολλούς ερευνητές στην εγκατάλειψη της βασικής έρευνας. Χωρίς όμως βασική έρευνα δεν μπορούν να υπάρξουν μετα-ανάλυσεις.

Οι επιπτώσεις της καθιέρωσης της μεθόδου της μετα-ανάλυσης στην προσπάθεια για τη διαπίστωση της εγκυρότητας των ψυχολογικών μετρήσεων άρχισαν να διαφαίνονται. Σε πολλά επίμαχα θέματα της ψυχολογίας έχουν γίνει τα τελευταία χρόνια μετα-ανάλυσεις. Σε πολλά επιστημονικά περιοδικά δε δημοσιεύουν πλέον μια και μοναδική έρευνα αλλά μόνο σειρά ερευνών στο ίδιο θέμα. Παράλληλα, χωρίς να αρνούνται τη βασική της αξία, έχουν αρχίσει οι διάφοροι προβληματισμοί για αυτή την ίδια τη μετα-ανάλυση. Οι προβληματισμοί αυτοί αναφέρονται κυρίως στις ειδικές μεθόδους που χρησιμοποιούνται, στο κατά πόσο οι δείκτες που με στατιστικό τρόπο συνάγονται για τον πληθυσμό είναι διογκωμένοι ή όχι, και στο ποσοστό της πιθανόν εσφαλμένης ανθρωπίνης κρίσης που υπεισέρχεται σε αυτή τη φαινομενικά αντικειμενική μέθοδο (π.χ., σχετικά με την επιλογή ή την απόρριψη των ερευνών που μετα-αναλύονται). Προσωπική άποψη του συγγραφέα του παρόντος άρθρου είναι ότι ένα βασικό θέμα σχετικά με τη μετα-ανάλυση είναι ότι πολλές έρευνες, ιδίως αυτές που δε βρίσκουν κάποια

στατιστικά σημαντική σχέση, δεν έχουν δημοσιευθεί, με αποτέλεσμα οι μετα-αναλύσεις να στηρίζονται σε ελλιπές δείγμα ερευνών. Αυτό πιθανό να διαγκώνει ή να μειώνει τεχνητά τους δείκτες που ανάγονται στον πληθυσμό. Το θέμα αυτό δεν έχει υπεισέλθει στο μέχρι τώρα προβληματισμό, γιατί βεβαίως κανείς δε γνωρίζει τον αριθμό των μη δημοσιευμένων ερευνών.

Βιβλιογραφία

- Hunter, J. E., & Hunter, R. F. (1984). Validity and ability of alternate predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J. E., & Schmidt, F. L. (1989). Meta-analysis: Facts and theories. In M. Smith & I. T. Robertson (Eds.), *Advances in selection and assessment*. Chichester: Wiley.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting errors and bias in research findings*. Newbury Park, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulative research findings across studies*. Beverly Hills, CA: Sage.
- Κάντας, Α. (1993). *Οργανωτική-βιομηχανική ψυχολογία*. Τόμ. 2: *Επιλογή και αξιολόγηση προσωπικού*. Αθήνα: Ελληνικά Γράμματα.
- Murphy, K. (1994). Meta-analysis and validity generalization. In N. Anderson & P. Herriot (Eds.), *Assessment and selection in organizations. First Update Supplement* (pp. 57-76). Chichester: Wiley.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.
- Smither, R. D. (1988). *The psychology of work and human performance*. New York: Harper & Row.

Problems associated with the validity of psychological measures

ARISTOTLE KANTAS
University of Patras, Greece

ABSTRACT

The validity of psychological measures is determined by the reliability of the predictor, on one hand, and the appropriateness of the criterion, on the other. Criterion relevance, criterion deficiency and criterion contamination are three aspects that have to be carefully examined during the choice of the criterion. Other factors like sampling error and range restriction can also hinder the determination of the true validity coefficient between the predictor and the criterion. Through the meta-analysis of existing research the true validity can be determined, across specific situations (validity generalization).

Keywords: Meta-analysis, validity, validity generalization.