

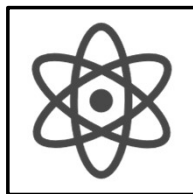
Βασικά χαρακτηριστικά της αξιολόγησης και των εξεταστικών δοκιμασιών με βάση την κλασική θεωρία

Γεώργιος Τσίτας

ΠΜΣ «ΠΟΛΙΤΙΚΗ, ΔΙΟΙΚΗΣΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ» (ΠΔΑΕ)

Αξιολόγηση στην Εκπαίδευση

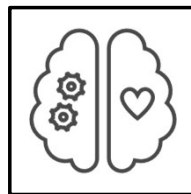
Βασικές έννοιες



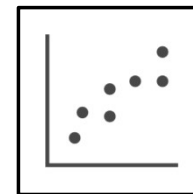
Εγκυρότητα



Αξιοπιστία



Αντικειμενικότητα



Διακριτικότητα



Δυσκολία



Πρακτικότητα

Πώς θεμελιώθηκε η σύγχρονη ψυχομετρία και συσχετίστηκε με την εκπαίδευση;

Το 1905 ο **Αλφρέντ Μπινέ** (Alfred Binet) και ο **Θεόδωρος Σιμόν** (Theodore Simon) ανέπτυξαν το πρώτο τεστ νοημοσύνης, γνωστό ως **Binet-Simon Scale**.

Το τεστ κατασκευάστηκε κατόπιν εντολής του γαλλικού Υπουργείου Παιδείας, να εντοπίσει παιδιά που χρειάζονταν ειδική εκπαίδευση.

Ο Μπινέ και ο Σιμόν επινόησαν μία σειρά από εργασίες που αξιολογούσαν τη μνήμη, την προσοχή, την κατανόηση, και τις λεκτικές δεξιότητες των παιδιών, προκειμένου να κατανοήσουν τις διαφορές στη νοητική ανάπτυξη.

Το έργο τους αποτέλεσε θεμέλιο για τη σύγχρονη ψυχομετρία και έθεσε τις βάσεις για τη δημιουργία μεταγενέστερων τεστ νοημοσύνης, όπως το **Stanford-Binet Intelligence Scale** (Lewis Terman ΗΠΑ).

Η ψυχομετρία αποτελεί ένα από τα κύρια πεδία της ψυχολογίας και περιλαμβάνει την κατασκευή και την αξιολόγηση τεστ για διάφορες νοητικές και ψυχολογικές ιδιότητες.

Έκτοτε...

με πρωτοβουλία διαφόρων επιστημόνων αναζητήθηκε ένας τρόπος κατασκευής εξεταστικών δοκιμασιών που να εξασφαλίζουν εγκυρότητα και αξιοπιστία.

Για να επιτευχθεί αυτό κρίνεται απαραίτητος ο έλεγχος των επιμέρους ερωτήσεων - ζητημάτων.

Ο έλεγχος αυτός ονομάζεται ανάλυση ερωτήσεων.

Στην πραγματικότητα, πρόκειται για έλεγχο των απαντήσεων που δίνονται στις ερωτήσεις.

Η κλασική θεωρία ανάγεται στον Spearman (1913). Στον αντίποδά της βρίσκεται η νεότερη θεωρία (IRT).

Σύμφωνα με την κλασική θεωρία:

- ▶ Κάθε μέτρηση είναι μια εμπειρική δειγματική εκτίμηση της πραγματικότητας που δεν αντικατοπτρίζει πάντοτε απόλυτα τις γνώσεις και τις ικανότητες των ανθρώπων.
- ▶ Η παρατηρούμενη επίδοση (X) προκύπτει από την εκτίμηση της πραγματικής επίδοσης (T), συνεκτιμώντας το θετικό ή αρνητικό σφάλμα (E) που εμπεριέχει, και που θεωρείται ίδιο για όλους τους εξεταζομένους (αυτό δεν σημαίνει ότι όλοι οι εξεταζόμενοι έχουν το ίδιο μέγεθος σφάλματος σε κάθε περίπτωση, αλλά ότι το σφάλμα είναι ανεξάρτητο και ομοιόμορφα κατανομημένο στον πληθυσμό).
- ▶ Έτσι, $X = T + E$
- ▶ Όσο πιο μικρό το E , τόσο πιο κοντά το X στο T .
- ▶ Αν μηδενίσουμε το E , τότε $X = T$.

Σε εξετάσεις μεγάλου εύρους, όπως είναι οι εξετάσεις με Τράπεζα Θεμάτων ή οι Πανελλαδικές εξετάσεις.

Δυστυχώς, στην Ελλάδα δεν υπάρχει κουλτούρα «στάθμισης» των θεμάτων, παρά μόνο εκ των υστέρων αποτίμησή τους.

Ειδικά στην Τράπεζα Θεμάτων, ο βαθμός δυσκολίας τους δεν ερείδεται σε επιστημονικά δεδομένα, γεγονός που καθιστά τη διαδικασία έωλη.

Υπάρχουν 1) Συλλογικές και 2) Ατομικές τράπεζες θεμάτων

1. Οι Συλλογικές εκπονούνται από επιτροπές ειδικών επιστημών, ή κρίνονται από αυτές όταν συντάσσονται από εκπαιδευτικούς.

2. Οι Ατομικές συλλέγονται από τους ίδιους τους εκπαιδευτικούς, ώστε να μην επαναλαμβάνονται σε σύντομα διαστήματα (2-3 έτη) τα ίδια θέματα και καθίστανται S.O.S.

Το μεγαλύτερο μέρος των ερωτήσεων θα πρέπει να έχουν δείκτη διακριτικότητας 0,5 και δυσκολίας (0,4-0,7), πρέπει, όμως, να περιλαμβάνονται και άλλες ερωτήσεις με μικρότερους ή υψηλότερους δείκτες.

Επίσης, προτείνεται να χρησιμοποιείται η IRT.

Σε κάθε περίπτωση, ο υπολογισμός των δεικτών απαιτεί στάθμιση, δηλαδή δαπάνη χρόνου και χρημάτων.

Πότε έχει νόημα να ασχολείται κάποιος με αυτά τα στοιχεία;

- Ας εξετάσουμε την ΕΒΕ «ελάχιστη βάση εισαγωγής».
- Πρόκειται για έναν μηχανισμό που καθορίζει (από το 2021) την κατώτατη βαθμολογία που απαιτείται για την εισαγωγή σε κάποιο πανεπιστημιακό τμήμα. Λειτουργεί ως φίλτρο που διασφαλίζει ότι οι υποψήφιοι που θα εισαχθούν στα ΑΕΙ πληρούν συγκεκριμένα ακαδημαϊκά κριτήρια.
- Πώς υπολογίζεται η ΕΒΕ
- Η ΕΒΕ δεν είναι σταθερή και δεν είναι ίδια για όλα τα τμήματα. Ορίζεται κάθε χρόνο για κάθε τμήμα χωριστά και υπολογίζεται ως ποσοστό επί του μέσου όρου των επιδόσεων των υποψηφίων που επιτεύχθηκαν στα τέσσερα πανελλαδικώς εξεταζόμενα μαθήματα σε κάθε επιστημονικό πεδίο. Η βάση αυτή ορίζεται από τα ίδια τα πανεπιστημιακά τμήματα, και κυμαίνεται μεταξύ του 0,8 και του 1,2 του γενικού μέσου όρου.

Παράδειγμα Υπολογισμού

- Αν ο μέσος όρος επιδόσεων των υποψηφίων σε ένα επιστημονικό πεδίο είναι 12/20, τότε:
- Ένα τμήμα που έχει ορίσει την ΕΒΕ στο 0,8 του μέσου όρου, θα έχει βάση 9,6 (12 x 0,8).
- Ένα άλλο τμήμα που έχει ορίσει την ΕΒΕ στο 1,2 του μέσου όρου, θα έχει βάση 14,4 (12 x 1,2).
- Τελικά, η ΕΒΕ λειτουργεί διορθωτικά ή όχι;

Εκτός από την Τράπεζα Θεμάτων, τι επίπτωση έχουν αυτοί οι δείκτες σε άλλες εξετάσεις π.χ. στις Πανελλαδικές;



Εγκυρότητα

- Η **εγκυρότητα** μιας εξέτασης στην εκπαίδευση αναφέρεται στο κατά πόσο η εξέταση μετράει όντως αυτό που προορίζεται να μετρήσει, ώστε να εξασφαλίζεται ότι τα αποτελέσματα της εξέτασης αντανακλούν με ακρίβεια τις δεξιότητες, τις γνώσεις ή τις ικανότητες των μαθητών στο συγκεκριμένο γνωστικό αντικείμενο.
- Η εγκυρότητα αναφέρεται στα αποτελέσματα και όχι γενικά στην εξέταση. Άρα, τα αποτελέσματα μίας εξέτασης παρουσιάζουν υψηλό ή χαμηλό βαθμό εγκυρότητας.

Εγκυρότητα Περιεχομένου (Content Validity)

Αναφέρεται στο κατά πόσο η εξέταση (δηλαδή οι ερωτήσεις ή οι εργασίες που περιλαμβάνει) είναι σύμφωνη με την **ύλη** και τους **στόχους** που είχαν τεθεί, καλύπτοντας όλα τα σημαντικά θέματα ή τις δεξιότητες που είναι απαραίτητο να αξιολογηθούν.

Για να επιτευχθεί η κάλυψη όλης της ύλης, προτείνεται η σύνταξη Πίνακα Προδιαγραφών, σαν αυτόν που ακολουθεί, ο οποίος περιλαμβάνει 10 θέματα. Με αυτόν τον τρόπο, η εξέταση είναι αντιπροσωπευτική όλης της ύλης και των στόχων.

Είδη Εγκυρότητας

Στόχοι	Ανάκληση	Εφαρμογή	Λύση Προβλήματος	Συνεισφορά στον βαθμό
Δυνατότητα πρόσθεσης	1 ερώτημα (10% βαθμού)	2 ερωτήματα (20% βαθμού)		30%
Δυνατότητα αφαίρεσης	1 ερώτημα (10% βαθμού)	2 ερωτήματα (20% βαθμού)		30%
Σύνθεση των δύο πράξεων σε πρόβλημα			2 ερωτήματα (40% βαθμού)	40%



Είδη Εγκυρότητας

Εγκυρότητα Εννοιολογικής κατασκευής (Construct Validity)

Αφορά το κατά πόσο η εξέταση μετράει την έννοια ή την ικανότητα που ισχυρίζεται ότι μετράει. Για παράδειγμα, αν μια εξέταση ισχυρίζεται ότι μετράει την κριτική σκέψη, θα πρέπει να περιέχει ερωτήσεις που όντως αξιολογούν αυτή τη δεξιότητα. Είναι πολύτιμη όταν η εξέταση έχει ως στόχο να αξιολογήσει μη χειροπιαστές έννοιες, όπως η δημιουργικότητα ή η κατανόηση.

Απαιτεί επιπλέον διαδικασίες:

- υπολογισμό συνάφειας με αντίστοιχα τεστ,
- διαφοροποίηση μεταξύ ομάδων που έχουν δεχθεί διαφορετικές παρεμβάσεις, π.χ. ενισχυτική διδασκαλία,
- ανάλυση παραγόντων κ.α.

Είδη Εγκυρότητας

Εγκυρότητα Κριτηρίου (Criterion-Related Validity)

Αξιολογεί την αποτελεσματικότητα της εξέτασης να προβλέπει ή να συσχετίζεται με ένα εξωτερικό κριτήριο.

Προγνωστική Εγκυρότητα (Predictive Validity): Δείχνει πόσο καλά τα αποτελέσματα της εξέτασης προβλέπουν μελλοντική απόδοση (π.χ. αν οι επιδόσεις στις Πανελλαδικές προβλέπουν επιτυχία στις σπουδές, κάτι που έχει επιτευχθεί στα Scholastic Assessment Test στις ΗΠΑ).

Συντρέχουσα Εγκυρότητα (Concurrent Validity): Δείχνει πόσο καλά οι επιδόσεις σε μία εξέταση συμφωνούν με τις επιδόσεις σε άλλες καθιερωμένες εξετάσεις για το ίδιο αντικείμενο την ίδια στιγμή.

Συχνά, για να αποτυπωθεί ο βαθμός συσχέτισης χρησιμοποιείται ο τύπος Spearman, r_{ho} που κυμαίνεται μεταξύ -1 και 1:

1 σημαίνει τέλεια θετική συσχέτιση (όσο αυξάνεται η μία μεταβλητή, αυξάνεται και η άλλη).

-1 σημαίνει τέλεια αρνητική συσχέτιση (όσο αυξάνεται η μία μεταβλητή, μειώνεται η άλλη).

0 σημαίνει ότι δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών.

Σχέση με την Εγκυρότητα Κριτηρίου

Περί Εγκυρότητας

Παράγοντες που επηρεάζουν την εγκυρότητα

Εσωτερικοί παράγοντες:

Ασάφειες, μοτίβα, υπόδειξη απάντησης, δύσκολο λεξιλόγιο, ερωτήσεις με πολύ μεγάλη ή πολύ μικρή δυσκολία, ανεπιτυχείς ερωτήσεις ως προς την ύλη ή τους στόχους, έκταση και χρόνος δοκιμίου (ελάχιστα ή τεράστια), διάταξη ή διασύνδεση ερωτήσεων...

Εξωτερικοί παράγοντες:

Συνθήκες χορήγησης (χρόνος, ακρίβεια στην εκφώνηση, βοήθεια), αντιγραφή, ορθή ή λανθασμένη διόρθωση κτλ.

Η αξιοπιστία (που ακολουθεί) είναι αναγκαία αλλά όχι επαρκής συνθήκη για την εγκυρότητα.

Μεταξύ εγκυρότητας και αξιοπιστίας η πρώτη είναι «σημαντικότερη». Αν βάλουμε 10 ίδια θέματα, θα πάρουμε ίδια αποτελέσματα (υψηλή αξιοπιστία), αλλά δεν θα έχουμε εικόνα όλης της ύλης.

Αξιοπιστία

Reliability

Η αξιοπιστία στην εκπαιδευτική αξιολόγηση αναφέρεται στον βαθμό στον οποίο ένα τεστ, μια αξιολόγηση ή ένα εργαλείο μέτρησης παράγει συνεπή και σταθερά αποτελέσματα. Ένα εργαλείο αξιολόγησης είναι αξιόπιστο όταν, κάτω από τις ίδιες συνθήκες, δίνει παρόμοια αποτελέσματα κάθε φορά που χρησιμοποιείται. Επομένως, η αξιοπιστία διασφαλίζει ότι τα αποτελέσματα της αξιολόγησης δεν επηρεάζονται από τυχαίους παράγοντες, π.χ. την ώρα της εξέτασης, λάθη στη μέτρηση ή ασάφειες στη διατύπωση των ερωτήσεων.

Η αξιοπιστία αναφέρεται στα αποτελέσματα της εξέτασης και όχι στην εξέταση καθεαυτή.

Η αξιοπιστία δεν εγγυάται την εγκυρότητα. Μπορεί τα αποτελέσματα να είναι αξιόπιστα αλλά όχι έγκυρα. Αν όμως δεν έχουμε αξιοπιστία, δεν έχουμε ούτε εγκυρότητα.

Η αξιοπιστία μπορεί να υπολογιστεί μόνο στατιστικά, όχι λογικά.

Το τυπικό σφάλμα μέτρησης είναι αντιστρόφως ανάλογο προς την αξιοπιστία. Θέλουμε χαμηλό τυπικό σφάλμα για να εξασφαλίσουμε μεγάλη αξιοπιστία.

Πώς συγκρίνουμε δύο μετρήσεις που δεν μπορεί να είναι ακριβώς ίδιες;

Μέθοδος επαναχορήγησης (Test-retest reliability): αναφέρεται στον βαθμό στον οποίο τα αποτελέσματα παραμένουν συνεπή όταν το ίδιο τεστ δίνεται σε ένα άτομο ή ομάδα σε διαφορετικές χρονικές στιγμές, δίνοντας έναν συντελεστή σταθερότητας (coefficient of stability).

Μέθοδος συνάφειας κριτών (Inter-rater reliability): Αφορά τη συνέπεια στις βαθμολογίες διαφορετικών αξιολογητών στις ίδιες απαντήσεις. Εάν οι κριτές δώσουν παρόμοιες βαθμολογίες σε ένα τεστ, τότε η αξιολόγηση θεωρείται αξιόπιστη. Ο συντελεστής συσχέτισης Pearson ή ο συντελεστής Spearman μπορεί να χρησιμοποιηθούν για να εκτιμήσουν τη συνάφεια μεταξύ των κριτών.

Μέθοδος διχοτόμησης των δοκιμίων (split-half method) που δίνει έναν συντελεστή εσωτερικής συνέπειας μεταξύ των ερωτήσεων του τεστ που έλαβαν μονό αριθμό και εκείνων που έλαβαν, για τις ανάγκες της συσχέτισης, ζυγό αριθμό. Κατόπιν ορίζεται ο δείκτης Spearman-Brown.

• **Η Μέθοδος των Kuder-Richardson** που δίνει συντελεστή εσωτερικής συνέπειας (coefficient of interval consistency) σε ερωτήσεις διχοτομικού χαρακτήρα.

Μέθοδος Cronbach alpha για ανοιχτού τύπου ερωτήσεις και τιμές από 0 έως 1, όπου: $\alpha \geq 0,9$: Εξαιρετική αξιοπιστία (πολύ υψηλή εσωτερική συνοχή), $0,8 \leq \alpha < 0,9$: Καλή αξιοπιστία, $0,7 \leq \alpha < 0,8$: Αποδεκτή αξιοπιστία, $0,6 \leq \alpha < 0,7$: Μέτρια αξιοπιστία (χρειάζεται βελτίωση), $\alpha < 0,6$: Χαμηλή αξιοπιστία (προβλήματα στην εσωτερική συνοχή).

Ποια η σημασία της επίδρασης του χρόνου, των γνώσεων που προστίθενται, του κινήτρου των μαθητών και της μνήμης; (όριο οι 12-15 μέρες).

Καθορισμός Αξιοπιστίας

Οι υπολογισμοί γίνονται με προγράμματα στατιστικής. Η ασφάλεια των δεικτών εξαρτάται και από παράγοντες όπως π.χ. ο αριθμός των ερωτήσεων ή οι αξιολογητές στις ερωτήσεις ανοιχτού τύπου.

Τι επηρεάζει την αξιοπιστία;

Η αξιοπιστία επηρεάζεται αρνητικά ή θετικά από ποικίλους παράγοντες:

Το πλήθος των ερωτημάτων. Όσο περισσότερα τα ερωτήματα, τόσο πιο αξιόπιστα τα αποτελέσματα (εφόσον όμως επαρκεί π.χ. ο χρόνος).

Εύρος αποτελεσμάτων – ετερογένεια συμμετεχόντων: όσο μεγαλύτερο το εύρος των αποτελεσμάτων και η ετερογένεια των εξεταζομένων, τόσο πιο αξιόπιστα τα αποτελέσματα.

- Αντικειμενικότητα βαθμολόγησης: όσο πιο αντικειμενική η βαθμολόγηση τόσο πιο αξιόπιστα τα αποτελέσματα, γι' αυτό και «υπερτερούν» οι κλειστού τύπου ερωτήσεις.
- Διακριτικότητα: όσο μεγαλύτερος ο δείκτης διακριτικότητας, τόσο πιο αξιόπιστα τα αποτελέσματα του τεστ (υπό την προϋπόθεση ότι οι ερωτήσεις είναι πολλές).

Η μεγάλη δυσκολία στα θέματα μπορεί να δώσει φαινομενικά αξιόπιστα αποτελέσματα, γιατί θα επαναλαμβάνονταν στο μέλλον, όμως δεν είναι όντως αξιόπιστα, επειδή π.χ. οι μαθητές/τριες απαντούν στην τύχη.

Δείκτης δυσκολίας των ερωτήσεων κλειστού τύπου

Χρήση του Δείκτη Δυσκολίας

Ο δείκτης αυτός βοηθά τους εκπαιδευτικούς να κατανοήσουν ποιες ερωτήσεις είναι πολύ εύκολες ή πολύ δύσκολες, ώστε να τις προσαρμόσουν κατάλληλα σε μελλοντικά τεστ, διατηρώντας έτσι την αξιολόγηση στο επιθυμητό επίπεδο δυσκολίας. Θέλουμε περισσότερες ερωτήσεις μέτριας δυσκολίας και λιγότερες μικρής ή μεγάλης.

Ας υποθέσουμε ότι ο παρακάτω πίνακας μάς δείχνει τον δείκτη δυσκολίας 5 διαφορετικών ερωτήσεων ενός διαγωνίσματος. Ποια από τις ερωτήσεις είναι η δυσκολότερη;

Ερώτηση	Δείκτης Δυσκολίας
1	0,1
2	0,3
3	0,5
4	0,7
5	0,9

Υπολογισμός δυσκολίας των ερωτήσεων κλειστού τύπου

Ο υπολογισμός βασίζεται στον αριθμό των μαθητών που απαντούν σωστά την ερώτηση.

Εκφράζεται με έναν τύπο που δίνει τιμές από 0 έως 1.

Ο τύπος για τον υπολογισμό του δείκτη δυσκολίας (Δ) είναι:

$$\Delta = \frac{N_c}{N}$$

όπου:

- N_c είναι ο αριθμός των μαθητών που απάντησαν σωστά στην ερώτηση.
- N είναι ο συνολικός αριθμός των μαθητών που απάντησαν την ερώτηση.

Ερμηνεία του Δείκτη Δυσκολίας

- 0,00 - 0,20: Πολύ δύσκολη ερώτηση (μόνο λίγοι μαθητές απαντούν σωστά).
- 0,21 - 0,40: Δύσκολη ερώτηση.
- 0,41 - 0,60: Μέτριας δυσκολίας ερώτηση.
- 0,61 - 0,80: Εύκολη ερώτηση.
- 0,81 - 1,00: Πολύ εύκολη ερώτηση (οι περισσότεροι μαθητές απαντούν σωστά).

ΠΡΟΣΟΧΗ!

Ένας δείκτης δυσκολίας κοντά στο 0 σημαίνει ότι η ερώτηση είναι πολύ δύσκολη, καθώς ελάχιστοι μαθητές την απάντησαν σωστά. Αντιθέτως, ένας δείκτης δυσκολίας κοντά στο 1 δείχνει ότι η ερώτηση είναι πολύ εύκολη, καθώς σχεδόν όλοι οι μαθητές την απάντησαν σωστά.

Υπολογισμός δυσκολίας των ερωτήσεων ανοιχτού τύπου

Ο υπολογισμός βασίζεται στον Μ.Ο. των βαθμών που πήραν οι μαθητές στη συγκεκριμένη ερώτηση αριθμό δια του του μέγιστου βαθμού που θα μπορούσαν να λάβουν σε αυτή την ερώτηση.

Εκφράζεται με έναν τύπο που δίνει τιμές από 0 έως 1.

Ο τύπος για τον υπολογισμό του δείκτη δυσκολίας (Δ) είναι:

$$\Delta = \frac{\text{Μ. ερ.}}{\text{Μ μεγ.}}$$

όπου:

- **Μ. ερ.** είναι ο Μ.Ο. των βαθμών που έλαβαν οι μαθητές στην ερώτηση.
- **Μ. μεγ.** είναι το άριστα της ερώτησης

π.χ.

$$\Delta = \frac{1}{10} = 0,1 \text{ (πολύ δύσκολη ερώτηση)}$$

Υπολογισμός δυσκολίας ολόκληρου του τεστ

Ο υπολογισμός βασίζεται στη διαίρεση του αθροίσματος των δεικτών δυσκολίας όλων των ερωτήσεων δια του πλήθους των ερωτήσεων.

Ο τύπος για τον υπολογισμό του συνολικού δείκτη δυσκολίας (Δ_{σ}) είναι:

$$\Delta_{\sigma} = \frac{\Sigma}{N}$$

όπου:

- Σ είναι το άθροισμα των Δ που έλαβαν οι μαθητές στα ερωτήματα.
- N το σύνολο των ερωτήσεων

π.χ. για τεστ 5 ερωτημάτων:

$$\Delta_{\sigma} = \frac{0,1+0,9+0,5+0,1+0,1}{5} = 0,34 \text{ (δύσκολο τεστ)}$$

Δείκτης διάκρισης

Ο δείκτης διάκρισης μιας ερώτησης δείχνει πόσο καλά η ερώτηση μπορεί να διαχωρίσει τους μαθητές με υψηλή απόδοση από αυτούς με χαμηλότερη απόδοση, βοηθώντας στον εντοπισμό των ερωτήσεων που είναι ικανές να διαχωρίσουν αποτελεσματικά τους μαθητές που κατέχουν την εξεταζόμενη γνώση από εκείνους που δεν την κατέχουν.

Διαδικασία Υπολογισμού

Διαχωρισμός των μαθητών σε δύο ομάδες: Με βάση τις συνολικές επιδόσεις τους στο τεστ, οι μαθητές χωρίζονται συνήθως στο ανώτερο και στο κατώτερο 27% (ή 30%, ανάλογα με την επιλογή) των επιδόσεων.

Αριθμός Σωστών Απαντήσεων: Μετράμε πόσοι μαθητές απάντησαν σωστά την ερώτηση στην ανώτερη και στην κατώτερη ομάδα.

Εφαρμογή του Τύπου: Υπολογίζεται ο δείκτης διάκρισης βάσει των σωστών απαντήσεων στις δύο ομάδες.

Υπολογισμός του Δείκτη Διάκρισης

Ο δείκτης διάκρισης (D) υπολογίζεται χρησιμοποιώντας τον παρακάτω τύπο:

$$D = \frac{O_{\alpha} - O_{\beta}}{N}$$

- O_{α} είναι ο αριθμός των μαθητών της ομάδας υψηλής επίδοσης που απάντησαν σωστά την ερώτηση.
- O_{β} είναι ο αριθμός των μαθητών της ομάδας χαμηλής επίδοσης που απάντησαν σωστά την ερώτηση.
- N είναι ο αριθμός των μαθητών σε κάθε ομάδα.

Ο δείκτης διάκρισης είναι ιδιαίτερα χρήσιμος για την αξιολόγηση της ποιότητας των ερωτήσεων, επιτρέποντας στους εκπαιδευτικούς να εντοπίσουν και να διορθώσουν ή να αφαιρέσουν ερωτήσεις που δεν επιτυγχάνουν τον στόχο τους στην εκπαιδευτική αξιολόγηση.

Δείκτης διάκρισης

Παράδειγμα

Δίδεται ο ακόλουθος πίνακας:

Ομάδα	Σύνολο μαθητών (N)	Σωστές Απαντήσεις (O)	Δείκτης Διάκρισης (D)
Ανώτερη (υψηλής επίδοσης)	10 (που έγραψαν γενικά πολύ καλά)	8 (απάντησαν σωστά και σε αυτή την ερώτηση)	
Μεσαία	17 (που έγραψαν μέτρια)	8 (απάντησαν σωστά σε αυτή την ερώτηση)	Δεν υπολογίζεται
Κατώτερη (χαμηλής επίδοσης)	10 (που δεν είχαν καλή επίδοση στο τεστ)	3 (απάντησαν σωστά και σε αυτή την ερώτηση)	
Υπολογισμός			$D = \frac{O_a - O_b}{N}$
Αποτέλεσμα			$D = \frac{8 - 3}{10} = 0,5$

Ανώτερη Ομάδα (Υψηλής Επίδοσης): Από τους 10 μαθητές σε αυτήν την ομάδα, οι 8 απάντησαν σωστά την ερώτηση.

Μεσαία Ομάδα Επίδοσης: Αυτή η ομάδα, που συνήθως αποτελείται από μαθητές με ενδιάμεση επίδοση, περιλαμβάνει 10 μαθητές, εκ των οποίων οι 5 απάντησαν σωστά. Οι απαντήσεις αυτής της ομάδας δεν επηρεάζουν τον υπολογισμό του δείκτη διάκρισης.

Κατώτερη Ομάδα (Χαμηλής Επίδοσης): Από τους 10 μαθητές στην κατώτερη ομάδα, οι 3 απάντησαν σωστά την ερώτηση.

Η **Μεσαία Ομάδα Επίδοσης** αναφέρεται για λόγους πληρότητας. Η συμμετοχή της στον υπολογισμό του δείκτη διάκρισης είναι μηδενική.

Ερμηνεία του Δείκτη Διάκρισης

0,40 και άνω: Η ερώτηση έχει υψηλή διακριτική ικανότητα

0,30 - 0,39: Η ερώτηση έχει καλή διακριτική ικανότητα.

0,20 - 0,29: Η ερώτηση έχει μέτρια διακριτική ικανότητα και μπορεί να χρειαστεί βελτίωση.

0,19 και κάτω: Η ερώτηση έχει χαμηλή ή αρνητική διακριτική ικανότητα, που σημαίνει ότι μπορεί να μην διαχωρίζει αποτελεσματικά τους μαθητές ή ακόμα και να οδηγεί σε συγχύσεις. Τέτοιες ερωτήσεις συνήθως χρειάζονται αναθεώρηση ή απομάκρυνση από το τεστ.

Δείκτης διάκρισης

Παράδειγμα σε ερωτήσεις ανοιχτού τύπου

Υπολογισμός του Δείκτη Διάκρισης

Ο δείκτης διάκρισης (D) υπολογίζεται χρησιμοποιώντας τον παρακάτω τύπο:

$$D = \frac{M_{\alpha} - M_{\beta}}{M_{\text{μεγ}}}$$

- M_{α} Μ.Ο. α' (υψηλόβαθμης ομάδας)
- M_{β} Μ.Ο. β' (χαμηλόβαθμης ομάδας)
- M μέγιστος βαθμός στην ερώτηση

Π.χ.

$$D = \frac{4,1 - 2,1}{5} = \frac{2}{5} = 0,4$$

Η ερώτηση έχει υψηλή διακριτική ικανότητα.

Σχέση δυσκολίας - διακριτικότητας

Συμπληρωματική Φύση: Ο δείκτης δυσκολίας μας λέει αν μια ερώτηση είναι εύκολη ή δύσκολη, ενώ ο δείκτης διακριτικότητας μάς δείχνει αν η ερώτηση είναι κατάλληλη για τη διάκριση μεταξύ μαθητών υψηλών και χαμηλών επιδόσεων. Μια καλή ερώτηση, ιδανικά, θα έχει έναν μεσαίο δείκτη δυσκολίας και υψηλό δείκτη διακριτικότητας.

Τι συμβαίνει λοιπόν με τις πολύ δύσκολες ή τις πολύ εύκολες ερωτήσεις;

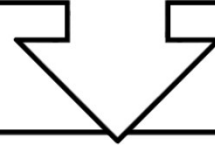
Εύκολες ερωτήσεις (με υψηλό δείκτη δυσκολίας) τείνουν να έχουν χαμηλότερη διακριτικότητα, γιατί οι περισσότεροι μαθητές τις απαντούν σωστά, άρα δεν επιτρέπουν εύκολα τον διαχωρισμό.

Δύσκολες ερωτήσεις (με χαμηλό δείκτη δυσκολίας) συχνά έχουν επίσης χαμηλή διακριτικότητα, καθώς λίγοι μαθητές τις απαντούν σωστά, χωρίς να διαχωρίζουν αποτελεσματικά ανάμεσα σε μαθητές υψηλών και χαμηλών επιδόσεων.

Μέτριες ερωτήσεις (με μέσο δείκτη δυσκολίας) συχνά επιτυγχάνουν υψηλότερη διακριτικότητα, καθώς απαντώνται σωστά από μαθητές με καλύτερη γνώση του θέματος και λάθος από άλλους.

Η **ιδανική ισορροπία** μεταξύ των δύο δεικτών βοηθά τους εκπαιδευτικούς να δημιουργούν αξιολογήσεις που είναι προκλητικές, αλλά δίκαιες, επιτρέποντας καλύτερη μέτρηση των πραγματικών διαφορών στις επιδόσεις των μαθητών.

Αντικειμενικά θεωρούνται τα αποτελέσματα μίας εξέτασης που ΔΕΝ επηρεάζονται από άλλους παράγοντες, π.χ. διαθέσεις, υποκειμενικές αντιλήψεις, στάσεις, συνθήκες, καταστάσεις ή άλλα προσωπικά χαρακτηριστικά των αξιολογητών.



Η ισοτιμία των αξιολογουμένων προϋποθέτει ομοιομορφία στην εξέταση και τη διόρθωση – βαθμολόγηση.

Ομοιομορφία εξέτασης: (χρόνος, συνθήκες, οδηγίες, διευκρινίσεις, κτλ.).

Ομοιομορφία βαθμολόγησης: συμφωνία για την εκτίμηση ορθού – λάθους, σαφή και ενιαία κριτήρια, ξεκάθαρο προσδιορισμό της βαρύτητας.

Αντικειμενικότητα

Πρακτικότητα και οικονομία

**Πρακτικότητα
-
χρηστικότητα**

Ευκολία προετοιμασίας εξέτασης - δημιουργίας εργαλείων ή μέσων

Κατανόηση και ανταπόκριση στη διαδικασία από πλευράς εξεταζομένων

Ευχέρεια διόρθωσης και ανακοίνωσης αποτελεσμάτων

Ευκολία εξασφάλισης κατάλληλων συνθηκών

**Κόστος -
οικονομία**

Απασχόληση - πληρωμή απαραίτητου εκπαιδευτικού προσωπικού

Ποσότητα χρόνου - όγκος εργασίας

Κόστος - δαπάνες διαδικασιών και διόρθωσης - βαθμολόγησης

Αδειοδοτήσεις, πρωτόκολλα, πνευματικά δικαιώματα

Αναστοχασμός

Ένα πράγμα
που έμαθα
σήμερα και που
δεν ήξερα έως
τώρα.

**Θα ήθελα να
σκεφτείτε:**

Ένα πράγμα
που χρειάζομαι
να μάθω
περισσότερα
γι' αυτό.

Ένα πράγμα
που έμαθα και
που με
εξέπληξε.

Ενδεικτική βιβλιογραφία

- ▶ Δημητρόπουλος, Ε. (2007). *Εκπαιδευτική Αξιολόγηση, η αξιολόγηση της εκπαίδευσης και του εκπαιδευτικού έργου, Α' μέρος*, Πέμπτη Έκδοση, εκδ. Γρηγόρη, Αθήνα
- ▶ Κασσωτάκης, Μ. (2013). *Η αξιολόγηση της επίδοσης των μαθητών*, 3^η αναθεωρημένη έκδοση, εκδ. Γρηγόρη, Αθήνα
- ▶ Κασσωτάκης, Μ. (2010). *Η αξιολόγηση της επίδοσης των μαθητών*, αναθεωρημένη έκδοση, εκδ. Γρηγόρη, Αθήνα
- ▶ Καψάλης, Α., & Χανιωτάκης, Ν. (2015). *Εκπαιδευτική Αξιολόγηση*, β' έκδοση, εκδ. Αφοί Κυριακίδη Α.Ε., Θεσσαλονίκη
- ▶ Κολιάδης, Ε. (2002). *Μέτρηση και αξιολόγηση της επίδοσης για τη διασφάλιση της επιτυχίας*, (Πρόγραμμα: Qual-Impact) Comenius 3.1, Αθήνα
- ▶ Κωνσταντίνου, Χ. (2007). *Η αξιολόγηση της επίδοσης του μαθητή ως παιδαγωγική λογική και σχολική πρακτική*, εκδ. Gutenberg, Αθήνα
- ▶ Παπαναστασίου, Κ. (2017). *Μέτρηση και Αξιολόγηση στην Εκπαίδευση*, Λευκωσία
- ▶ Παρασκευόπουλος Ι. (1990) *Στατιστική εφαρμοσμένη στις επιστήμες της συμπεριφοράς*, Τόμος Α', Περιγραφική Στατιστική