

ΕΡΕΥΝΗΤΙΚΗ ΕΡΓΑΣΙΑ ORIGINAL PAPER

Βελτιστοποίηση των εξετάσεων πολλαπλής επιλογής με τη χρήση ειδικού λογισμικού

ΣΚΟΠΟΣ Το σύστημα εξετάσεων με ερωτήσεις πολλαπλής επιλογής χρησιμοποιείται συχνά για την αξιολόγηση των φοιτητών. Το περιεχόμενο των ερωτήσεων επιλέγεται κατά κανόνα από το διδάσκοντα με βάση την κρίση και την εμπειρία του. Σκοπός της παρούσας εργασίας ήταν η ανάπτυξη ενός προγράμματος αξιολόγησης των ερωτήσεων πολλαπλής επιλογής που δίνονται στους φοιτητές της Ιατρικής Σχολής. **ΥΛΙΚΟ-ΜΕΘΟΔΟΣ** Αναπτύχθηκε ειδικό λογισμικό πρόγραμμα, με τη βοήθεια του οποίου δημιουργήθηκαν βάσεις ερωτήσεων, όπου μεταφέρθηκε το σύνολο των ερωτήσεων πολλαπλής επιλογής που χρησιμοποιούνται για την εξέταση των φοιτητών Ιατρικής. Για όλες τις ερωτήσεις της βάσης υπολογίστηκαν οι συντελεστές ευκολίας και διακριτικότητας. Ο συντελεστής διακριτικότητας αντανάκλα τη διακριτική ικανότητα της ερώτησης, δηλαδή το κατά πόσο έχει υψηλό ποσοστό επιτυχίας στους φοιτητές με υψηλή βαθμολογία. Αξιολογήθηκαν 95 ερωτήσεις πολλαπλής επιλογής, που είχαν χρησιμοποιηθεί σε πληθυσμό 176 ± 76 (72–441) φοιτητών. Για την αξιολόγηση των απαντήσεων χρησιμοποιήθηκε ειδικός σαρωτής και λογισμικό. **ΑΠΟΤΕΛΕΣΜΑΤΑ** Οι ερωτήσεις που αξιολογήθηκαν είχαν απαντηθεί σωστά από τους φοιτητές σε ποσοστό (ποσοστό επιτυχίας) $53 \pm 22\%$. Ως συνιστώμενης δυσκολίας αξιολογήθηκαν 15/95 ερωτήσεις (16%), ενώ 13/95 (14%) θεωρήθηκαν ως «υψηλής δυσκολίας-μη αποδεκτής» και 21/95 (22%) ως «υψηλής ευκολίας-μη αποδεκτής». Άριστη διακριτικότητα είχαν 38/95 ερωτήσεις (40%), ενώ διαπιστώθηκαν 16/95 ερωτήσεις με κακή διακριτικότητα (17%). **ΣΥΜΠΕΡΑΣΜΑΤΑ** Η επιλογή των ερωτήσεων πολλαπλής επιλογής με μόνο κριτήριο την κρίση του διδάσκοντα φαίνεται ότι δεν επαρκεί για τη δημιουργία ενός κατά το δυνατόν αντικειμενικού συστήματος εξετάσεων. Η χρησιμοποίηση ειδικού λογισμικού είναι ένας εύκολος και ασφαλής τρόπος για τη βελτίωση της ποιότητας των ερωτήσεων πολλαπλής επιλογής.

Η διαδικασία της εκπαίδευσης των φοιτητών σε ένα γνωστικό αντικείμενο περιλαμβάνει, εκτός των παραδόσεων και της πρακτικής άσκησης, και την εξασφάλιση ενός συστήματος εξέτασης που θα μπορεί να αξιολογεί την επάρκεια των φοιτητών στο υπό εξέταση αντικείμενο.¹⁻³ Η ταξινόμηση του Bloom επιχείρησε να ομαδοποιήσει τους στόχους της εκπαίδευσης και χρησιμοποιείται ως πλαίσιο για την οργάνωση του προγράμματος σπουδών και των εξετάσεων. Παρότι οι εκπαιδευτές συνήθως καταβάλλουν σημαντική προσπάθεια στην προετοιμασία των παραδόσεων, εντούτοις συχνά δεν δίνουν αντίστοιχη βαρύτητα στην προετοιμασία των εξετάσεων, σπάνια δε έχουν εκπαιδευτεί στην τέχνη της εξέτασης.⁴

Οι ερωτήσεις πολλαπλής επιλογής αποτελούν ένα δημοφιλή τρόπο εξέτασης, που έχει επικρατήσει λόγω ορισμένων πλεονεκτημάτων. Επιτρέπουν την εξέταση μεγάλου αριθμού

εξεταζομένων και μεγάλου μέρους της εξεταστέας ύλης,^{5,6} είναι αντικειμενικές, βαθμολογούνται εύκολα και γρήγορα.⁶ Μειονεκτήματα της εξεταστικής αυτής μεθόδου είναι ότι ενθαρρύνουν την αντιγραφή, μπορεί να είναι δυσνόητες ή παραπλανητικές⁶ και αδυνατούν να εξετάσουν ανώτερες γνωστικές λειτουργίες, όπως η συνθετική σκέψη.^{6,7} Τέλος, ένα σημαντικό μειονέκτημα των ερωτήσεων πολλαπλής επιλογής είναι ότι η δημιουργία τους είναι δύσκολη και χρονοβόρα για τους διδάσκοντες.^{1,6}

Η δημιουργία ερωτήσεων πολλαπλής επιλογής διέπεται από ορισμένες αρχές,^{1,6} που στοχεύουν στη δημιουργία εξετάσεων που θα μπορούν να αξιολογήσουν σωστά τους φοιτητές. Μελέτες επιχείρησαν να αξιολογήσουν την ποιότητα των ερωτήσεων πολλαπλής επιλογής προερχόμενων από έτοιμες βάσεις ερωτήσεων^{5,7} ή διαμορφωμένων από έμπειρους διδάσκοντες.^{4,5,8,9} Διαπίστωσαν σημαντικό

ΑΡΧΕΙΑ ΕΛΛΗΝΙΚΗΣ ΙΑΤΡΙΚΗΣ 2008, 25(6):781–785
ARCHIVES OF HELLENIC MEDICINE 2008, 25(6):781–785

Σ. Νανάς,
Β. Γεροβασίλη,
Σ. Πουλάκη,
Α. Μπούχλα,
Ε. Τριποδάκη,
Θ. Λούκας,
Χ. Ρούτση,
Χ. Ρούσσο

Α΄ Πανεπιστημιακή Κλινική
Εντατικής Θεραπείας, Νοσοκομείο
«Ευαγγελισμός», Εθνικό και
Καποδιστριακό Πανεπιστήμιο Αθηνών,
Αθήνα

Optimization of multiple choice
examinations with the use of
specifically designed software

Abstract at the end of the article

Λέξεις ευρετηρίου

Αξιολόγηση
Ερωτήσεις πολλαπλής επιλογής
Ιατρική εκπαίδευση

Υποβλήθηκε 30.10.2007
Εγκρίθηκε 12.11.2007

αριθμό ατελειών (item writing flaws, IWF), με αντίστοιχες επιπτώσεις στην επίδοση των φοιτητών.

Στόχος της παρούσας μελέτης ήταν η ανάπτυξη ενός λογισμικού που θα μπορούσε να χρησιμοποιηθεί για την αξιολόγηση και τη βελτιστοποίηση των ερωτήσεων πολλαπλής επιλογής που δίνονται σε προπτυχιακά μαθήματα στους φοιτητές Ιατρικής.

ΥΛΙΚΟ ΚΑΙ ΜΕΘΟΔΟΣ

Λειτουργίες αναπτυχθέντος λογισμικού

Αναπτύχθηκε ειδικό λογισμικό πρόγραμμα (Quicktesting, ANOVA consulting), με τη βοήθεια του οποίου δημιουργήθηκαν βάσεις ερωτήσεων. Στις δημιουργηθείσες βάσεις μεταφέρθηκε το σύνολο των ερωτήσεων πολλαπλής επιλογής που χρησιμοποιούνται για την εξέταση των φοιτητών Ιατρικής στο μάθημα της εντατικής Θεραπείας, ομαδοποιημένων κατά κεφάλαιο. Για κάθε ερώτηση της βάσης περιέχονται πληροφορίες για το ιστορικό της ερώτησης, δηλαδή το δείγμα φοιτητών στο οποίο έχει χρησιμοποιηθεί, το ποσοστό των φοιτητών που την έχει απαντήσει σωστά (ποσοστό επιτυχίας) και το πότε χρησιμοποιήθηκε για τελευταία φορά η ερώτηση σε εξετάσεις. Οι πληροφορίες αυτές είναι δυνατόν να αξιοποιηθούν για την αξιολόγηση των ερωτήσεων πολλαπλής επιλογής.

Οι ερωτήσεις της βάσης μπορούν να ταξινομηθούν με βάση το κεφάλαιο στο οποίο ανήκουν, την ημερομηνία τελευταίας χρήσης τους και το ποσοστό επιτυχίας τους. Επίσης, είναι δυνατόν να επιλεγούν ερωτήσεις με βάση το βαθμό δυσκολίας ή και διακριτικότητάς τους. Ο βαθμός δυσκολίας και διακριτικότητας υπολογίζεται αυτόματα από το λογισμικό για κάθε ερώτηση της βάσης.

Στις δυνατότητες του λογισμικού περιλαμβάνονται επίσης η δημιουργία διαγωνισμάτων και η αναδιάταξη των ερωτήσεων και των απαντήσεων στο διαγώνισμα. Με αυτόν τον τρόπο προκύπτουν εναλλακτικές εκδοχές του αρχικού διαγωνίσματος που μπορούν κατόπιν να ομαδοποιηθούν, ώστε η διόρθωση να γίνεται αυτόματα για όλες.

Η διόρθωση των διαγωνισμάτων γίνεται με τη βοήθεια ειδικού σαρωτή (scanner), με δυνατότητα άμεσης διόρθωσής του και εξαγωγής των αποτελεσμάτων και των στατιστικών στοιχείων της εξέτασης εν γένει και των ερωτήσεων ξεχωριστά.

Αξιολόγηση των ερωτήσεων πολλαπλής επιλογής

Για την αξιολόγηση των ερωτήσεων πολλαπλής επιλογής χρησιμοποιούνται οι αυτόματα υπολογιζόμενοι από το λογισμικό συντελεστές δυσκολίας και διακριτικότητας, οι οποίοι χρησιμοποιούνται ευρέως για την ποιοτική αξιολόγηση των ερωτήσεων πολλαπλής επιλογής.^{1,8-11}

Με τη βοήθεια του ειδικού σαρωτή βαθμολογείται κάθε εξεταζόμενος ξεχωριστά. Ακολούθως, ταξινομούνται οι εξετασθέντες με

βάση το ποσοστό επιτυχίας τους στην εξέταση και καθορίζεται η ομάδα των φοιτητών υψηλής (25%) και χαμηλής (25%) απόδοσης. Οι δύο αυτές ομάδες χρησιμοποιούνται για τον υπολογισμό των δεικτών ευκολίας και διακριτικότητας κάθε ερώτησης πολλαπλής επιλογής.

Ο δείκτης δυσκολίας αντανakλά το ποσοστό επιτυχίας στην ερώτηση και παίρνει τιμές από 0 έως 100. Όσο μεγαλύτερος είναι ο δείκτης δυσκολίας τόσο πιο εύκολη είναι η ερώτηση, πρόκειται δηλαδή ουσιαστικά για δείκτη ευκολίας.¹ Ο δείκτης δυσκολίας υπολογίζεται από την εξίσωση: συντελεστής δυσκολίας= $(X+Ψ) \times 100/n$, όπου X είναι ο αριθμός των σωστών απαντήσεων της ομάδας υψηλής βαθμολογίας, $Ψ$ είναι ο αριθμός σωστών απαντήσεων της ομάδας χαμηλής βαθμολογίας και n είναι το άθροισμα των εξετασθέντων των δύο ομάδων.

Ο δείκτης διακριτικότητας αντανakλά την ικανότητα που έχει η ερώτηση να διακρίνει τους «υψηλούς» από τους «χαμηλούς» σε απόδοση φοιτητές. Ο δείκτης διακριτικότητας παίρνει τιμές από -1 ως +1. Όσο προσεγγίζει την τιμή +1, τόσο καλύτερη διακριτική ικανότητα έχει η ερώτηση.¹ Ο δείκτης διακριτικότητας υπολογίζεται από την εξίσωση: συντελεστής διακριτικότητας= $2 \times (X-Ψ)/n$, όπου X , $Ψ$ και n όπως ανωτέρω.

ΑΠΟΤΕΛΕΣΜΑΤΑ

Χρησιμοποιήθηκε δείγμα 95 ερωτήσεων πολλαπλής επιλογής, οι οποίες είχαν χρησιμοποιηθεί στο παρελθόν σε δείγμα 176 ± 76 (72–441) πεμπτοετών φοιτητών Ιατρικής.

Οι ερωτήσεις που χορηγήθηκαν είχαν ποσοστό επιτυχίας $53 \pm 22\%$. Με βάση το συντελεστή δυσκολίας, οι ερωτήσεις ομαδοποιήθηκαν σε πέντε ομάδες, όπως φαίνεται στον πίνακα 1, και με βάση το συντελεστή διακριτικότητας ομαδοποιήθηκαν σε τέσσερις ομάδες, όπως φαίνεται στον πίνακα 2.

Ως συνιστώμενης δυσκολίας αξιολογήθηκαν 15/95 ερωτήσεις (16%), ενώ 13/95 (14%) θεωρήθηκαν ως «υψηλής δυσκολίας-μη αποδεκτής» και 21/95 (22%) ως «υψηλής ευκολίας-μη αποδεκτής» (εικ. 1). Άριστη διακριτικότητα

Πίνακας 1. Ομαδοποίηση των 95 ερωτήσεων πολλαπλής επιλογής, που αξιολογήθηκαν με βάση το δείκτη δυσκολίας σε πέντε ομάδες.

Ομάδα	Δείκτης δυσκολίας
Υψηλής ευκολίας-μη αποδεκτής	≥ 70
Αποδεκτής ευκολίας	60–69
Συνιστώμενης δυσκολίας	50–59
Αποδεκτής δυσκολίας	30–49
Υψηλής δυσκολίας-μη αποδεκτής	< 30

Ο δείκτης δυσκολίας αντανakλά το ποσοστό επιτυχίας στην ερώτηση και παίρνει τιμές από 0–100

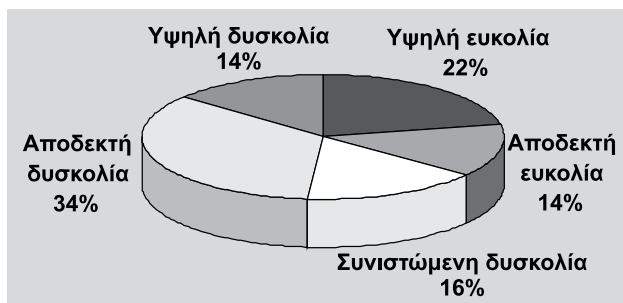
Πίνακας 2. Ομαδοποίηση των 95 ερωτήσεων πολλαπλής επιλογής, που αξιολογήθηκαν με βάση το δείκτη διακριτικότητας σε τέσσερις ομάδες.

Ομάδα	Δείκτης διακριτικότητας
Άριστη διακριτικότητα	>0,35
Καλή διακριτικότητα	0,25–0,34
Μέτρια διακριτικότητα	0,15–0,24
Κακή διακριτικότητα	<0,15

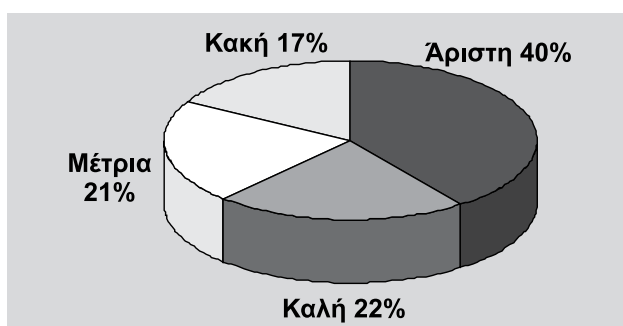
Ο δείκτης διακριτικότητας αντανακλά την ικανότητα που έχει η ερώτηση να διακρίνει τους «υψηλούς» από τους «χαμηλούς» σε απόδοση φοιτητές και παίρνει τιμές από -1 ως +1.

είχαν 38/95 ερωτήσεις (40%), ενώ διαπιστώθηκαν 16/95 ερωτήσεις με κακή διακριτικότητα (17%) (εικ. 2).

Οι ερωτήσεις συνιστώμενης δυσκολίας (15/95) είχαν καλή (5/7) και άριστη (9/7) διακριτικότητα, ενώ υπήρχε και μία ερώτηση κακής διακριτικότητας. Από τις ερωτήσεις «υψηλής δυσκολίας-μη αποδεκτής» (13/95), διαπιστώθηκε ότι 3/13 (23%) είχαν κακή διακριτικότητα και 6/13 (46%) είχαν μέτρια διακριτικότητα. Αντίστοιχα, από τις ερωτήσεις «υψηλής ευκολίας-μη αποδεκτής» (21/95), διαπιστώθηκε ότι 5/21 (24%) είχαν κακή και 7/21 (33%) είχαν μέτρια διακριτικότητα.



Εικόνα 1. Βαθμός δυσκολίας των 95 ερωτήσεων πολλαπλής επιλογής, που αξιολογήθηκαν από την ηλεκτρονική βάση ερωτήσεων.



Εικόνα 2. Βαθμός διακριτικότητας των 95 ερωτήσεων πολλαπλής επιλογής, που αξιολογήθηκαν από την ηλεκτρονική βάση ερωτήσεων.

Από τις ερωτήσεις «υψηλής δυσκολίας-μη αποδεκτής» (13/95), διαπιστώθηκε ότι 3 ερωτήσεις (3/13, 23%) είχαν λανθασμένη σωστή απάντηση ή δύο σωστές απαντήσεις. Διαπιστώθηκε επίσης ότι σε 3 ερωτήσεις (3/13, 23%) αναφερόταν το όνομα του φαρμακευτικού προϊόντος με την εμπορική του ονομασία.

Από τις ερωτήσεις κακής διακριτικότητας (16/95), διαπιστώθηκε ότι 5 ερωτήσεις (5/16, 31%) είχαν δύο τουλάχιστον απαντήσεις πολύ συναφείς και άλλες 5 ερωτήσεις (5/16, 31%) περιείχαν τουλάχιστον μία παραβίαση των κανόνων δημιουργίας ερωτήσεων πολλαπλής επιλογής (IWF).

ΣΥΖΗΤΗΣΗ

Το βασικό εύρημα που προκύπτει από τη μελέτη μας είναι ότι η ανάπτυξη ενός λογισμικού προγράμματος για την αξιολόγηση των ερωτήσεων πολλαπλής επιλογής οδήγησε στη διαπίστωση ύπαρξης ατελειών στις ερωτήσεις πολλαπλής επιλογής που διαθέτουμε. Εντούτοις, οι ερωτήσεις που αξιολογήθηκαν είχαν κριθεί κατάλληλες από τους διδάσκοντες του μαθήματος με βάση την κρίση και την εμπειρία τους. Διαπιστώθηκε, αρχικά, η ύπαρξη ερωτήσεων ακραίας δυσκολίας ή και κακής διακριτικότητας, η περαιτέρω διερεύνηση των οποίων βοήθησε στον εντοπισμό σφαλμάτων. Από τη βιβλιογραφική αναζήτηση δεν διαπιστώθηκε η καταγραφή λογισμικού προγράμματος αντίστοιχων δυνατοτήτων.

Η αξιολόγηση των ερωτήσεων με το λογισμικό αυτό παρέχει ορισμένα πλεονεκτήματα. Καθιστά δυνατή την ταχεία αξιολόγηση των ερωτήσεων, αφού η διαδικασία είναι αυτοματοποιημένη. Κατά συνέπεια, είναι ευκολότερος ο εντοπισμός σφαλμάτων και η συναξιολόγηση της ποιότητας των ερωτήσεων πολλαπλής επιλογής. Επίσης, είναι δυνατόν η αξιολόγηση να γίνει πριν από την ανακοίνωση των αποτελεσμάτων και, εφόσον εντοπιστούν σφάλματα, μπορούν αυτά να διορθωθούν και κατόπιν να ανακοινωθούν τα αποτελέσματα.

Υπάρχουν ορισμένοι λόγοι που ερμηνεύουν την ύπαρξη ερωτήσεων ακραίας δυσκολίας ή και κακής διακριτικότητας. Μια πιθανή ερμηνεία είναι ότι οι συγκεκριμένες ερωτήσεις περιέχουν κάποιο σφάλμα, όπως δύο σωστές απαντήσεις. Πιθανόν επίσης να είναι δυσνόητες ή διφορούμενες. Η ύπαρξη σφαλμάτων σε ερωτήσεις πολλαπλής επιλογής αποτελεί σύνηθες φαινόμενο,¹² ωστόσο οι μελέτες που έχουν αξιολογήσει την ύπαρξη σφαλμάτων σε ερωτήσεις πολλαπλής επιλογής που δίνονται σε φοιτητές Ιατρικής είναι περιορισμένες.^{4,7} Σε μια μελέτη αξιολογήθηκε η ποιότητα των ερωτήσεων πολλαπλής επιλογής ως χαμηλή,⁴ ενώ σε μια άλλη διαπιστώθηκε ότι 46% των ερωτήσεων

πολλαπλής επιλογής περιείχαν τουλάχιστον μία ατέλεια (IWF).⁷ Το ποσοστό των ερωτήσεων πολλαπλής επιλογής που περιέχουν τουλάχιστον μία ατέλεια βρέθηκε να είναι 75% σε μια μελέτη που αφορούσε σε φοιτητές λογιστικής.¹³ Στη δική μας μελέτη, η αναζήτηση των ερωτήσεων κακής διακριτικότητας διευκόλυνε τον εντοπισμό λανθασμένων ερωτήσεων, όπως ερωτήσεων χωρίς κάποια σωστή απάντηση ή με δύο συναφείς απαντήσεις, και τη διόρθωσή τους, ή ερωτήσεων που δεν πληρούσαν τους κανόνες για τη δημιουργία ερωτήσεων πολλαπλής επιλογής.¹⁴

Η σημασία της ύπαρξης ερωτήσεων πολλαπλής επιλογής που περιέχουν σφάλματα είναι ότι μπορεί να επηρεάσει την απόδοση των φοιτητών. Σε μια μελέτη, η αφαίρεση των ερωτήσεων πολλαπλής επιλογής που διαπιστώθηκε ότι περιέχουν σφάλματα θα επέτρεπε σε ποσοστό 10–15% των φοιτητών Ιατρικής που απέτυχαν να περάσουν τη βάση.⁹ Η ανάπτυξη ενός λογισμικού με τη δυνατότητα αυτόματου υπολογισμού των συντελεστών ευκολίας και διακριτικότητας διευκολύνει τον εντοπισμό τέτοιων ερωτήσεων, με σκοπό τη διόρθωσή τους και την καλύτερη αξιολόγηση των φοιτητών.

Ένα από τα σημαντικά μειονεκτήματα των ερωτήσεων πολλαπλής επιλογής είναι ότι ενθαρρύνουν την αντιγραφή. Η δυνατότητα δημιουργίας διαγωνισμάτων με αναδιάταξη των ερωτήσεων και των απαντήσεων μπορεί να συμβάλλει στον περιορισμό της αντιγραφής, βελτιώνοντας έτσι την ποιότητα των εξετάσεων. Η διαδικασία αυτή δεν επιβαρύνει περαιτέρω τους διδάσκοντες, καθότι τα εναλλακτικά διαγωνίσματα ομαδοποιούνται και η διόρθωση γίνεται ταυτόχρονα για όλες τις εναλλακτικές διατάξεις.

Συμπερασματικά, η ανάπτυξη ενός ειδικού λογισμικού για τη βελτιστοποίηση των εξετάσεων με τη χρήση ερωτήσεων πολλαπλής επιλογής έδωσε τη δυνατότητα να αξιολογηθούν οι ερωτήσεις πολλαπλής επιλογής, με βάση τη δυσκολία και τη διακριτική τους ικανότητα. Εντοπίστηκαν έτσι ερωτήσεις άριστης διακριτικής ικανότητας και συνιστώμενης δυσκολίας, αλλά και ερωτήσεις ακραίας δυσκολίας ή ευκολίας και κακής διακριτικότητας. Η περαιτέρω αξιολόγηση των ερωτήσεων αυτών θα επιτρέψει τον εντοπισμό λαθών αλλά και την καλύτερη αξιολόγηση του περιεχομένου των ερωτήσεων, με τελικό σκοπό την καλύτερη αξιολόγηση των φοιτητών.

ABSTRACT

Optimization of multiple choice examinations with the use of specifically designed software

S. NANAS, V. GEROVASILIS, S. POULAKI, A. BOUHLA, E. TRIPODAKI, T. LOUKAS, C. ROUTSI, C. ROUSSOS

1st Department of Critical Care Medicine, "Evangelismos" Hospital, National and Kapodestrian University of Athens, Athens, Greece

Archives of Hellenic Medicine 2008, 25(6):781–785

OBJECTIVE Multiple choice examinations are often used to assess student achievement. Question's content is chosen mainly by the tutor according to his/her judgment and experience. This project aimed to develop a program for evaluation of multiple choice questions used for testing medical students. **METHOD** Software was developed which enabled the creation of a question base into which all multiple choice questions used to examine medical students were transferred. For each question the difficulty and discrimination indices were calculated. The discrimination index represents the discrete faculty of each question that is whether it has a high rate of success within the students with high grades. Evaluation was made of 95 multiple choice questions used in a population of 176±76 (72–441) students; a special scanner and special software were used. **RESULTS** Questions evaluated were correctly answered in a percentage (success rate) of 53±22%. Of all the questions, 15/95 were considered to be of recommended difficulty (16%) while 13/95 (14%) were considered to be of "high difficulty-not acceptable" and 21/95 (22%) of "high facility-not acceptable". Of most excellent discrimination were 38/95 questions (40%) while 16/95 (17%) were considered to be of low discrimination. **CONCLUSIONS** Choosing multiple choice questions according to the tutor's judgment only, is not sufficient to create an objective evaluation system of students' knowledge. The use of specifically designed software is an easy and safe way of improving the quality of multiple choice questions.

Key words: Evaluation, Medical education, Multiple choice questions

Βιβλιογραφία

1. ΒΑΣΙΛΟΠΟΥΛΟΣ Δ. *Αρχές και μέθοδοι της ιατρικής εκπαίδευσης*. Εκδόσεις Σμίλη, Αθήνα, 1998
2. ANDERSON LW, KRATHWOHL DR. *A taxonomy for learning, teaching and assessing*. Pearson Education, New York, 2001
3. MARZANO RJ. *Classroom assessing and grading*. Association for supervision curriculum development, Alexandria, Virginia, USA, 2006
4. JOZEFOWICZ RF, KOEPPEL BM, CASE S, GALBRAITH R, SWANSON D, GLEW RH. The quality of in-house medical school examinations. *Acad Med* 2002, 77:156–161
5. TARRANT M, KNIERIM A, HAYES SK, WARE J. The frequency of item writing flaws in multiple choice questions used in high stakes nursing assessments. *Nurse Educ Today* 2006, 26:662–671
6. FARLEY JK. The multiple choice test. Writing the questions. *Nurse Educ* 1989, 14:10–39
7. MASTERS JC, HULSMAYER BS, PIKE ME, LEICHTY K, MILLER MT, VERST AL. Assessment of multiple choice questions in selected test banks accompanying text books used in nursing education. *J Nurs Educ* 2001, 40:25–32
8. DOWNING SM. Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? *Acad Med* 2002, 77(Suppl 10):S103–S104
9. DOWNING SM. The effects of violating standard item writing principles on tests and students. The consequences of using flawed test items on achievement examinations in medical education. *Adv Health Sci Educ* 2005, 10:133–143
10. De SANTIS M, McKEAN TA. Efficient validation of teaching and learning using multiple choice exams. *Adv Physiol Educ* 2003, 27:3–14
11. TOLLEFSON N, TRIPP A. Are complex multiple choice options more difficult and discriminating than conventional multiple choice options? *J Nurs Educ* 1985, 24:92–98
12. STAGNARO-GREEN AS, DOWNING SM. Use of flawed multiple-choice items by the New England Journal of Medicine for continuing medical education. *Med Teach* 2006, 28:566–568
13. HANSEN JD, DEXTER L. Quality multiple-choice test questions: Item-writing guidelines and an analysis of auditing test banks. *Journal of Education for Business* 1997, 73:94–97
14. BLOOM BS. *Taxonomy of educational objectives. Handbook 1: The cognitive domain*. Longman, London, 1956

Corresponding author:

S. Nanas, 1st Department of Critical Care Medicine, "Evangelismos" Hospital, National and Kapodestrian University of Athens, 45–47 Ipsilandou street, GR-106 75 Athens, Greece
e-mail: snanas@cc.uoa.gr